JET AND DARK SUBSTRUCTURE: RISE OF THE MACHINES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mathieu Cliche

August 2015

JET AND DARK SUBSTRUCTURE:

RISE OF THE MACHINES

Mathieu Cliche, Ph.D.

Cornell University 2015

ABSTRACT

We currently live in the golden age of experimental particle physics. The Large Hadron Collider (LHC) finally discovered the Higgs particle and may soon allow us to probe TeV-scale new physics. Moreover, dark matter indirect detection experiments may have found annihilation products in our galaxy while the direct detection experiments keep improving their upper bounds on dark matter interactions with ordinary matter. That being said, the current efforts to discover any physics beyond the standard model have failed. We thus need to come up with better models and better tools to probe new physics. This thesis which collects the author's PhD research papers is an effort in this direction.

We begin by showing how singular dark matter self-interaction can significantly enhance dark matter annihilation in today's Universe. Then we explore a concrete dark matter model where a particle called the dilaton mediate the interactions between the dark sector and the standard model. Finally we devise a new method to discern heavy boosted particles from QCD jets at the LHC.

BIOGRAPHICAL SKETCH

Mathieu Cliche grew up in Quebec city, Canada. It was in his last year of high school that Mathieu started to think about becoming a physicist after doing a school project on cosmology. That idea stuck with him throughout *CEGEP*[1], where he became a very self-motivated student.

Mathieu then earned a Bachelor of Science in physics at L'Universite Laval. During these years he participated in research on ferro-fluid mirrors at Laval university and accelerator physics at TRIUMF. After reading a book on quantum computing, Mathieu decided to pursue a Master's in quantum information at the University of Waterloo with Achim Kempf. However, after taking a class on quantum field theory Mathieu wanted to know more about particle physics so he decided to pursue this subject at Cornell University.

He thus pursued a Ph.D. at Cornell under Prof. Csaba Csaki who taught him about supersymmetry and extra-dimensions. Near the end of his PhD, Mathieu got interested in data science and started participating in machine learning competitions. After receiving his physics Ph.D., he will therefore pursue a data scientist career in New York city.

[1] In Quebec, high school is 5 years, then there is 2 years of *CEGEP* and then university where a B.Sc takes 3 years.

ACKNOWLEDGEMENTS

Over the years I have been fortunate enough to surround myself with really wonderful people. I owe the completion of this thesis to many of them.

- My advisor, Csaba Csaki, was a great teacher and mentor. He managed to give me the right amount of guidance and he was always very supportive of my occasional change of interests.
- Maxim Perelstein, Yuval Grossman, Lawrence Gibbons and Julia Thom also gave me invaluable teachings in particle physics.
- Flip Tanedo, Brando Bellazini, Seung Lee, Kfir Blum, Leandro Almeida and Mihailo Backovic were incredible collaborators who helped me with the grunt work associated with research.
- Katerina, Liz, Kacey and Deb helped me greatly with administrative duties.
- My friends: Bibhushan, Cody, David, Flip, Francois, Frank, Gui, Hussam, Jack, Javier, JD, Jeff, John, JP, Judit, Marco, Mario, Mike, Naresh, Nicolas, Paul, Riccardo, Steve, and the others I'm forgetting too, your moral support and companionship was invaluable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In this thesis we explore two different areas of particle physics, namely dark matter and collider searches. With dark matter, the goal is to expand the standard model to include one or more particles which would explain certain gravitational anomalies. This poses certain challenges since any extensions of the standard model must be consistent with current experimental data. Moreover, since dark matter is dark we need to be particularly ingenious to find ways to "see" it. Theoretical physicists see these challenges as an opportunity not only to solve the standard model's shortcomings but also to explain away certain astrophysical anomalies. With collider searches, our goal is to develop new tools to extract as much information as possible from the data generated by the LHC in order to constraint or probe new physics.

We therefore begin this chapter with a brief introduction to dark matter. Our goal is to present the historical context and current state of affairs before moving to original research in Chapter 2 and 3. We also discuss some topics which should render these later chapters more accessible, such as Sommerfeld enhancement and conformal symmetry breaking. Then, since in Chapter 4 we discuss a new tool for collider searches which rely heavily on machine learning, we introduce the topic here. This introduction is meant for particle physicists who may not have any prior knowledge of machine learning. We introduce a handful of algorithms and then discuss how some of them are used in particle physics applications.

1.1 Dark Matter

The first hint of dark matter came in 1933 with F. Zwicky who observed the radial velocity dispersion of galaxies in the Coma cluster [3]. Based on his observations he concluded that the mass-to-light ratio was larger in that cluster by two orders of magnitude than the same ratio in the solar neighborhood. A few years later, in 1939, astronomers found that the outer region of the Andromeda galaxy was rotating much faster than what would be predicted by Newtonian dynamics given the matter distribution based on luminous matter [4]. This has since become one of the most powerful argument for the existence of dark matter. A different solution would be to modify Newton's laws at large scales but this program has since been dropped because it was not able to explain the other pieces of evidence in favor of dark matter. One of these pieces comes from gravitational lensing caused by dark matter, that is when dark matter gravitationally affects the path of light between an astrophysical object and the earth. A classic example of the use of gravitational lensing happened in 2006 when astronomers observed the collision of two clusters of galaxies [5]. Using X-rays to locate the luminous collisional matter and gravitational lensing to locate the total gravitational mass, it was concluded that the luminous matter lagged behind the total mass as one would expect from weakly-interacting dark matter.

Cosmology with the direct measurement of the cosmic microwave background (CMB) spectrum also indicates the presence of dark matter. Indeed, the measurement of the fluctuations in the CMB can infer the relative densities of baryonic and non-baryonic matter in the Universe, from which one can conclude that most of the matter in the universe is made of non-baryonic dark matter [6]. In 2013 Planck's measurements inferred that ordinary matter makes 5% of the Universe while dark matter makes 27% of the Universe [7].

Our understanding of dark matter can therefore be summarized as follows. We know

precisely the relic density of dark matter in the Universe and its density distribution throughout the Universe is often very well known from gravitational lensing measurements. However we know very little about what dark matter is actually made of in terms of particles. This section will attempt to shed some light on this topic. We start by reviewing the theoretical and experimental status of the search for the identity of dark matter. We then discuss the possibility that dark matter self-interact and we present an overview of the previous work done in this direction. We end with a discussion of a particle called the dilaton which could play a role in dark matter self-interaction.

1.1.1 State of the field

Theoretical status

Our most hopeful attempt at discovering the identity dark matter comes from the weakly interacting massive particle (WIMP) miracle. To discuss this miracle we first need to compute the relic density of dark matter [8, 9]. We will assume that dark matter is stable and interacts weakly with the Standard model, both crucial assumptions for the WIMP miracle. This allows us to assume that in the early Universe dark matter is in thermal equilibrium with the rest of the matter, that is the production of dark matter particles from the thermal bath is equal to its annihilation rate. Now if the temperature of the Universe had adiabatically lowered to its present day value, the present day relic abundance would be thermally suppressed by $e^{-m_\chi/T_{\text{present}}}$ where m_χ is the dark matter mass and T_{present} is the present day temperature. This suppression factor would effectively mean that there is no dark matter left in the present Universe. However, the Universe does not cool down adiabatically because it is expanding. In fact, when the annihilation rate becomes much smaller than the expansion rate, the dark matter density freezes. This essentially means that when the Universe is expanding fast enough the dark matter particles can no longer

meet to annihilate. This process is quantified by the Boltzmann equation

$$\frac{dn_\chi}{dt} = -3Hn_\chi + \langle\sigma v\rangle \left[n_{\chi\text{EQ}}^2 - n_\chi^2\right] \tag{1.1}$$

where n_χ is the dark matter particle number density, $n_{\chi\text{EQ}}$ is the thermal equilibrium number density, H is the Hubble parameter of the Universe and $\langle\sigma v\rangle$ is the thermally averaged velocity-weighted annihilation cross section. We can solve this equation by assuming that the early Universe is radiation dominated and that $\sigma v \approx a + bv$ where a and b are parameters which depend on the dark matter model. The first assumption allows us to use the Hubble parameter $H \propto T^2$ to switch from the time variable t to the temperature variable T. The second assumption allows us to approximate the solution to get the relic abundance $\Omega_\chi h^2 = m_\chi n_\chi/\rho_{\text{crit}}$ where $\rho_{\text{crit}} = 3H^2/8\pi G$. We get

$$x_F = \ln\left(\frac{5}{4}\sqrt{\frac{45}{8}}\frac{g}{2\pi^3}\frac{M_{\text{Pl}} m_\chi (a + 6b/x_F)}{\sqrt{g^*}\sqrt{x_F}}\right) \tag{1.2}$$

$$\Omega_\chi h^2 \approx \frac{1.07\times 10^9}{\text{GeV} M_{\text{Pl}}\sqrt{g^*}}\frac{x_F}{a + 3(b - a/4)/x_F} \tag{1.3}$$

where $x_F = m_\chi/T_F$ and T_F is the freeze-out temperature. In addition g is the number of degrees of freedom of the dark matter particle and g^* is the effective number of relativistic degrees of freedom in thermal equilibrium during dark matter freeze-out. Plugging a few numbers in this equation allows us to approximate it further as

$$\Omega_\chi h^2 \approx 0.1\left(\frac{\alpha^2\langle v\rangle//(100\text{GeV})^2}{\langle\sigma v\rangle}\right). \tag{1.4}$$

Note that the measured relic abundance is $\Omega_\chi h^2 = 0.1199 \pm 0.0027$ [10]. As a consequence, we see from this equation that a weakly interacting particle with a mass around the weak scale gives the correct relic abundance; this is the WIMP miracle. Because of the hierarchy problem, we already have good reasons to believe that there are new particles around the weak scale that we have not yet observed. Therefore, this fortunate coincidence seems to indicate that the identity of dark matter is intrinsically related to the solution of the hierarchy problem.

The most popular solution to the hierarchy problem is Supersymmetry [11]. With Supersymmetry, all the Standard model particles have a superpartner that has a spin which differ by $1/2$. This solves the hierarchy problem since the one loop contribution to the Higgs mass parameter from a Standard model particle differs by a relative sign from its superpartner. If the mass of the superpartners were equal to their Standard model counterparts, their contribution to the Higgs mass would vanish. Since we want these contributions to be moderate in size, this requires the mass of the superpartners to be not too much heavier than the weak scale. So far none of these superpartners have been discovered. When Supersymmetry comes packed in with R-parity then we have a natural dark matter candidate. Indeed, to avoid having operators that can induce proton decay one often adds to Supersymmetry a discrete $\mathcal{Z}_2$ symmetry called R-parity. Standard model particles are even under R-parity while their superpartners are odd. The lightest superpartner (LSP) cannot decay to Standard model particles and is therefore stable. If the LSP is electromagnetically neutral then it is a perfect dark matter candidate. One possibility for the LSP is the lightest neutralino which is a combination of the bino (superpartner of the $U(1)$ gauge field), neutral wino (superpartner of the W gauge field), and neutral Higgsinos (superpartner of the Higgs boson).

Another popular possibility is that dark matter is composite. This first came up in models where the Higgs is also composite [12]. Indeed, note that light scalars like pions in QCD are perfectly natural since they are not elementary but instead bound states. These light scalars are pseudo-Goldstone bosons originating from the spontaneous breaking of a global symmetry by the dynamics of a strong sector. If the Higgs bosons is similarly composite then there is no hierarchy problem since its mass does not receive corrections above the compositeness scale f, just like pions do not receive Planckian loop corrections. A natural dark matter candidate can arise in these models when the breaking of the global symmetry gives an additional pseudo-Goldstone gauge singlet. Dark matter then usually

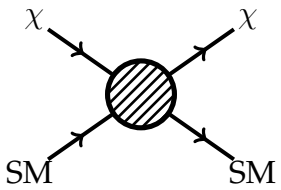

Figure 1.1: Diagram of interest for direct detection.

interacts with the Standard model via Higgs interactions. Later we will explore a concrete dark matter model from this category where the broken symmetry is conformal invariance.

Experimental status

Let us now briefly discuss the various ways one might hope to probe dark matter experimentally, either through direct detection, indirect detection or with collider searches.

The idea of direct detection is to study the elastic scattering of dark matter particles in our galaxy with the nucleus of shielded detectors [13]. The Feynman diagram of interest is illustrated on Fig. (1.1). The main challenge with direct detection is background reduction. Recoil events are rare so it is important that the various sources of background are as suppressed as possible. The main sources of background come from gamma and beta radiation from the surrounding area and the detector itself, from fast neutron coming from the cosmic ray shower and from solar neutrinos. To reduce this background these experiments are conducted deep underground. So far direct detection experiments have not seen convincingly any dark matter collisions. Fig. (1.2) summarizes the most recent results from the various experiments which are expressed as an upper bound on the direct detection cross section.

With indirect detection, the goal is to measure the products of dark matter annihila-

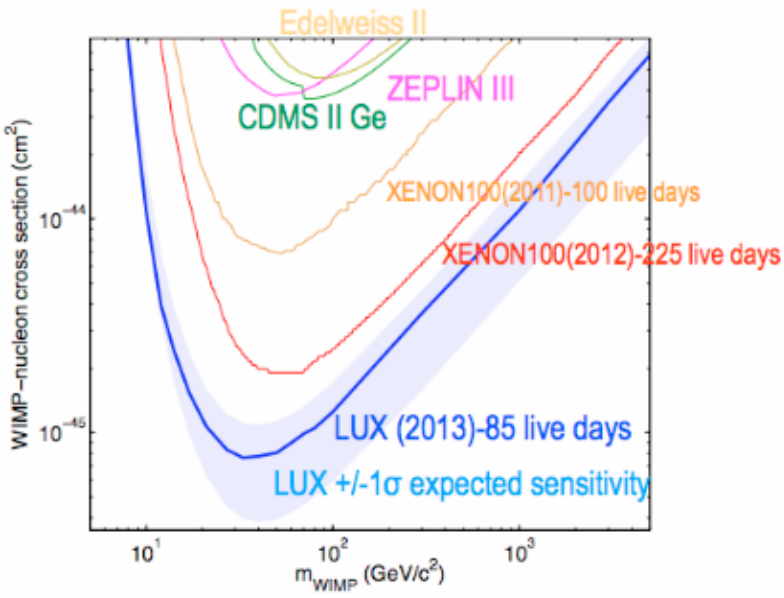

Figure 1.2: Direct detection experimental results [1].

tion [14]. Indeed, the same process which is responsible for dark matter annihilation in the early Universe, which is illustrated on Fig. (1.3), can still happen today in regions of space where dark matter is sufficiently dense. These processes could cause an excess of gamma rays, anti-protons or positron in those high dark matter density regions. Some experiments have observed excess in these signals; the Fermi gamma-ray space telescope has observed an excess in gamma radiation coming from the center of the Milky way at the 130 GeV line [15] while the PAMELA experiment observed an excess in the positron flux [16]. Note however that indirect detection is the type of detection which is the most sensitive to astrophysical uncertainties since we not only have to account for all the possible sources of background but also for the way the signal propagates from its source to our detectors.

The third way to detect dark matter is via collider searches. As illustrated on Fig. (1.4), this can happen at the LHC when protons collide and produce dark matter. Since dark matter interacts very weakly with ordinary matter it cannot be detected in the detectors,

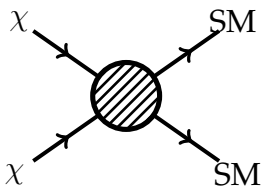

Figure 1.3: Diagram of interest for indirect detection.

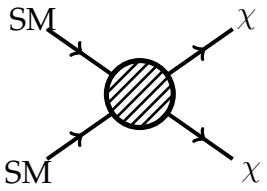

Figure 1.4: Diagram of interest for collider searches.

and since dark matter is stable or at least long-lived it does not decay to produce a visible signature of its own. Thus, if dark matter is produced at the LHC it will manifest itself as an excess of missing energy. This has not been observed yet but note that this approach has the advantage of being completely free of any astrophysical uncertainties.

1.1.2 Self-interaction

The first motivation for dark matter self-interaction came from the indirect detection anomalies. For the excess observed to be explained by WIMP annihilations, it requires the annihilation cross section to be significantly greater than what is required to get the right relic abundance. However, there is a mechanism called Sommerfeld enhancement which can boost the annihilation cross section at small velocities without changing it during the early Universe. This Sommerfeld enhancement is an effect which can happen when dark matter self-interact before annihilating. This possible explanation for indirect detection anomalies made dark matter self-interaction a trendy topic around 2009, however its mo-

tivation has since evolved more towards small scale astrophysical anomalies [17]. One of these anomalies, known as the core vs cusp problem, is that observations of dwarf galaxies show dark matter distribution with cores while numerical simulations of cold collisionless dark matter show a cusp. Nevertheless, self-interactions can transfer the energy from the outer halo to the central region which then forms a core. Indeed, recent simulations show that all the small scale anomalies can be solved if dark matter self-interact with a cross section of $\sigma_{SI}/m_\chi \sim 0.1 - 10\text{cm}^2/g$ at dwarf galaxies scales. While this cross section is significantly larger than the typical weak scale cross section it is easy to obtain with a light dark force. Since light mediators often already exist in dark matter models this is encouraging.

To get an idea of the physics behind self-interacting dark matter we now briefly review how the Sommerfeld enhancement and the elastic cross section are calculated [18]. From the particle physics side these two quantities completely determine the phenomenology of self-interacting dark matter. The idea behind Sommerfeld enhancement is that the attractive force between the two dark matter particles makes it more likely that the particles meet and annihilate, thus boosting the effective annihilation cross section. This is intrinsically a non-relativistic problem so we use non-relativistic quantum mechanics to solve it. The two dark matter particles interacting form a two-body system which we can reduce to a one-body central force problem in the center of mass frame. In that frame the annihilation happens at the origin where there is a point-like interaction Hamiltonian $H_{\text{ann}} = U_{\text{ann}}\delta^3(\mathbf{r})$. The force between the two dark matter particles is described by a non-relativistic potential $V(r)$. Note that $V(r)$ is the only particle physics input to this problem. In other words, once a dark matter model is cooked up, one can simply extract the non-relativistic potential $V(r)$ and then forget about the intricacies of the model from that point on. Before annihilating, the wave-function of this system can be described by a particle moving in the z direction $\psi(\mathbf{x})^{(0)}_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{x}}$. The rate of annihilation should be

proportional to how often the particle is seen at the origin, that is to $|\psi(\mathbf{0})|^2$. The dynamics of the wave-function is determined by solving the Schrödinger equation

$$-\frac{1}{2m_r}\nabla^2\psi_k + V(r)\psi_k = \frac{k^2}{2m_r}\psi_k \tag{1.5}$$

where $m_r = m_\chi/2$ is the reduced mass. Notice that the interaction $V(r)$ will affect the value of the wave-function at the origin and therefore will impact the effective annihilation rate, this is the core of the Sommerfeld enhancement. To solve the Schrödinger equation we expand the wave-function in terms of partial waves

$$\psi_k = \sum_l A_l P_l(\cos(\theta)) R_{kl}(r) \tag{1.6}$$

where $P_l(\cos(\theta))$ are Legendre polynomials. Plugging this partial wave decomposition into the Schrödinger equation gives the following differential equation for $R_{kl}(r)$

$$-\frac{-1}{2m_r}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}R_{kl}\right) + \left(\frac{l(l+1)}{r^2} + V(r)\right)R_{kl} = \frac{k^2}{2m_r}R_{kl}. \tag{1.7}$$

To solve this we need two boundary conditions. If the potential does not blow up faster than $1/r$ near the origin, than regularity of R_{kl} requires that $R_{kl} \sim r^l$ near the origin. Furthermore, we impose at infinity the following normalization

$$R_{kl}(r \to \infty) \to \frac{1}{r}\sin(kr - \frac{l\pi}{2} + \delta_l) \tag{1.8}$$

where δ_l are phase shifts used in the computation of the elastic cross section. Indeed, once the Schrödinger equation is solved with our two boundary conditions, we can extract the Sommerfeld enhancement from the wave-function at the origin and extract the elastic cross section from the wave-function at infinity. We get that the differential elastic cross section is

$$\frac{d\sigma_{\text{elas}}}{d\Omega} = \frac{1}{k^2}\left|\sum_{l=0}^{\infty}(2l+1)e^{2i\delta_l}P_l(\cos(\theta))\sin(\delta_l)\right|^2 \tag{1.9}$$

while the Sommerfeld enhancement for an $l = 0$ wave is

$$S_{k,l=0} = \left|\frac{\psi(0)}{\psi^{(0)}(0)}\right|^2 = \left|\frac{R_{k,l=0}(0)}{k}\right|^2 \tag{1.10}$$

such that $\sigma_{\text{ann}} = \sigma^{(0)}_{\text{ann}} S_{k,l=0}$ where $\sigma^{(0)}_{\text{ann}}$ is the bare annihilation cross section. In principle if our annihilation cross section contains contribution from higher partial waves these needs to be Sommerfeld enhanced accordingly as well. A complete derivation of the Sommerfeld enhancement for higher partial waves can be found in Appendix B.

As a sanity check it is easy to see that if we have $V(r) = 0$ then $R_{k,l=0} = \sin(kr)/r$ such that $\delta_l = 0$, $\frac{d\sigma_{\text{elas}}}{d\Omega} = 0$ and similarly we get that $S_{k,l=0} = 1$. For generic potentials one has to solve the Schrödinger equation numerically. However for the Coulomb potential $V(r) = -\frac{\alpha}{r}$, which often happens when the self-interaction is mediated by a massless boson, then we have the following analytical formula for the Sommerfeld enhancement of an $l = 0$ wave

$$S_{k,l=0} = \frac{\frac{2\pi\alpha}{v}}{1 - e^{-\frac{2\pi\alpha}{v}}} \tag{1.11}$$

where v is the relative velocity between the two dark matter particles. Note that as expected when $v \to \infty$ we get $S_{k,l=0} \to 1$, while at small velocities we get $S_{k,l=0} \approx \frac{2\pi\alpha}{v}$ which can be a significant boost.

1.1.3 The dilaton

As we previously discussed, one interesting possibility is that dark matter is composite. This possibility first came up in models where the Higgs is a pseudo-Goldstone boson originating from the spontaneous breaking of a global symmetry. When the broken global symmetry is sufficiently large it is natural to have an additional gauge singlet which can be a dark matter candidate. In chapter 3, we will explore a different scenario where instead of a natural dark matter candidate we have a natural dark matter self-interaction. The self-interaction is mediated by a new particle called the dilaton, which is simply the Goldstone boson of scale invariance spontaneously broken. In addition to being a natural

self-interaction mediator, the dilaton could also be a natural mediator between dark matter and the Standard model. We therefore introduce in this section the dilaton and discuss its couplings to other fields. We mostly follow [19, 20].

Under a scale transformation $x \to x' = e^{-\lambda}x$ operators transform as $\mathcal{O}(x) \to \mathcal{O}'(x) = e^{\lambda\Delta}\mathcal{O}(e^{\lambda}x)$ where Δ is the dimension of the operator $\mathcal{O}(x)$ which includes both classical and quantum effects. Similarly the action transforms as

$$S = \sum_i \int d^4x g_i \mathcal{O}_i(x) \to S' = \sum_i \int d^4x e^{\lambda(\Delta_i - 4)} g_i \mathcal{O}_i(x). \tag{1.12}$$

We thus see that if all operators have $\Delta_i = 4$ then the action is invariant under scale transformation. We now assume that scale invariance is broken spontaneously at a scale f, which means there is a Goldstone boson associated with scale transformation. This Goldstone boson $\widetilde{\sigma}(x)$ is called the dilaton and it transforms inhomogeneously under a scale transformation $\widetilde{\sigma}(x) \to \widetilde{\sigma}'(x) = \widetilde{\sigma}(e^{\lambda}x) + \lambda f$. In order to obtain the low energy effective theory we can build a conformal compensator $\Phi(x)$,

$$\Phi(x) = f e^{\widetilde{\sigma}(x)/f}. \tag{1.13}$$

This conformal compensator transform as $\Phi(x) \to e^{\lambda}\Phi(e^{\lambda}x)$ and has a vev $\langle\Phi(x)\rangle = f$. It therefore acts as a spurion field, that is we can insert powers of $\Phi(x)/f$ in any Lagrangian to make it scale invariant.

Let us now briefly discuss the coupling of the dilaton to other fields. To do so we make the assumption that there is two different sectors. We have a spontaneously broken conformal sector which we refer to as the composite sector, and a sector which we refer to as the elementary sector which explicitly breaks the scale invariance through small couplings to the composite sector. In the composite sector the high energy field content has a Lagrangian which is scale invariant (if we neglect small explicit breaking terms). However, below the scale $\Lambda \sim 4\pi f$, the low energy field content is different and may need

powers of Φ to preserve scale invariance

$$\mathcal{L}^{\text{IR}}_{\text{CFT}} = \sum_i c_i \mathcal{O}^{\text{IR}}_i \Phi^{m_i}. \tag{1.14}$$

To preserve scale invariance we require that $m_i = 4 - \Delta^{\text{IR}}_i$. As a concrete example, suppose that in the low energy effective theory we have a non-interacting composite scalar χ. Its kinetic term is already scale invariant and therefore does not need any powers of Φ, however its mass term does. We thus get

$$\mathcal{L}^{\text{IR}}_{\text{CFT}} = \partial_\mu \chi \partial^\mu \chi - \frac{1}{2} m_\chi^2 \chi^2 \Phi^2 \tag{1.15}$$

where we fixed $m_i = 2$ for the mass term. From this Lagrangian it would then be straightforward to extract the interaction between χ and $\widetilde{\sigma}$ by expanding Φ in powers of $\widetilde{\sigma}$. Note that if we only include the linear term in $\widetilde{\sigma}$ then Eq. (1.14) is equivalent to saying that the dilaton couples linearly to the trace of the energy momentum tensor.

When the elementary fields are weakly coupled to composite fields then they also have interactions with the dilaton. To see how this works it is best to work with a concrete example. Suppose we have two elementary fermions ϕ_L and ψ_R which interact weakly with the composite operators Θ_L and Θ_R of the UV theory

$$\mathcal{L}^{UV}_{\text{int}} = y_L \psi_L \Theta_R + y_R \psi_R \Theta_L + h.c. \tag{1.16}$$

This interaction breaks scale invariance so we assign spurius dimensions to the coupling constants, $[y_L] = 4 - \Delta^{UV}_{\psi_L} - \Delta^{UV}_{\Theta_R}$ and $[y_R] = 4 - \Delta^{UV}_{\psi_R} - \Delta^{UV}_{\Theta_L}$. In the low energy effective theory we can integrate out the massive composite fields and get

$$\mathcal{L}^{IR}_{\text{int}} = -M y_L y_R \psi_L \psi_R \Phi^m + h.c. \tag{1.17}$$

The value of m is determined by requiring scale invariance

$$m = 4 - ([y_L] + [y_R]) - \Delta^{IR}_{\psi_L} - \Delta^{IR}_{\psi_R} \approx \Delta^{UV}_{\Theta_L} + \Delta^{UV}_{\Theta_R} - 4 \tag{1.18}$$

where we neglected the perturbative anomalous dimensions of the elementary fermions. This can also be written as $m = 1 + \gamma_L + \gamma_R$ since $\Delta^{UV}_{\Theta_L} = 5/2 + \gamma_L$ and $\Delta^{UV}_{\Theta_R} = 5/2 + \gamma_R$ where γ is the anomalous dimension. At linear order in the dilaton field this gives

$$\mathcal{L}^{IR}_{\text{int}} = -m_\psi \psi_L \psi_R - m_\psi \left(1 + \gamma_L + \gamma_R\right) \psi_L \psi_R \frac{\widetilde{\sigma}}{f} + h.c. \tag{1.19}$$

where $m_\psi = M y_L y_R$.

Similarly, if an elementary gauge field A_μ is coupled to a composite current $\mathcal{J}^\mu$ in the UV

$$\mathcal{L}^{UV}_{\text{int}} = -\frac{1}{4 g^2_{UV}} F_{\mu\nu} F^{\mu\nu} + A_\mu J^\mu \tag{1.20}$$

then we must assign spurius dimension to the coupling constant $[g_{UV}] = \Delta^{UV}_A - 1$. Below the symmetry breaking scale f we then get an effective theory

$$\mathcal{L}^{IR}_{\text{int}} = -\frac{1}{4 g^2} F_{\mu\nu} F^{\mu\nu} \Phi^m. \tag{1.21}$$

For scale invariance we require that $m = 4 - 2\left(1 + \Delta^{IR}_A\right) + 2[g]$, and since $[g] = [g_{UV}]$ this leads to

$$m = 2\left(\Delta^{UV}_A - \Delta^{IR}_A\right) = 2\left(\frac{\beta_{IR}}{g} - \frac{\beta_{UV}}{g}\right) \tag{1.22}$$

where β is the beta function. This means that the coupling between the dilaton and our elementary gauge field is proportional to the difference in the beta functions of g above and below f. Note that if our gauge field is instead entirely composite then we have $\Delta^{UV}_A = 1$ such that $\beta_{UV} = 0$ and this reduces to the well known case of the trace anomaly.

1.1.4 Summary

We started with a historical review of the evidence for dark matter. We then made the case for a weakly interacting massive particle with the calculation of the relic abundance

through thermal freeze-out. We also argued that supersymmetry and compositeness can generate natural WIMP candidates. We followed this by briefly discussing the various ways one may hope to experimentally probe dark matter. One of these approaches, namely indirect detection, may be greatly affected by an additional force in the dark sector. We thus discussed how dark matter self-interaction may affect the annihilation cross section relevant for indirect detection. We ended with a brief review of the dilaton, a particle which could mediate the dark sector and the Standard model.

1.2 Machine Learning

We now switch gear and start discussing a completely different topic, namely machine learning. Machine learning is arguably the most exciting field in the modern age of technology. Its goal is to have computers adapt their behavior over time so that they get better at the task they are assigned. This is an ambitious goal that is inherently multidisciplinary, taking ideas from statistics, neuroscience, mathematics and computer science. It is often divided into the following sub-categories

- **Supervised learning:** The data given to the computer is labeled. This means that the computer takes in a training data set and attempts to generalize the label assignment to any input data.
- **Unsupervised learning:** The data given to the computer is not labeled. This means that the computer's goal is to find similarities between the inputs and categorized them and any other inputs accordingly.
- **Reinforcement learning:** No data is given to the computer, but instead the computer can interact with a dynamical environment. The computer is being told when the answer is wrong, but not how to correct it. This means it has to explore the possible

solutions until it figures out the correct one.

Reinforcement learning is used for example when one wants to teach a computer how to play chess, but it is otherwise rarely used. Supervised and unsupervised learning are the most common types of machine learning.

Let us now discuss supervised learning in more details. Suppose we have some data points $\mathbf{x}_i$ of dimension m which are correlated with targets $\mathbf{t}_i$ of dimension n. The components of $\mathbf{x}$ are called features and the space of these data points is called the feature space. We can think of the data points $\mathbf{x}_i$ as the inputs to a machine. With these inputs the goal of supervised learning is to come up with a machine which produces outputs $\mathbf{y}_i$ in the target space such that $\mathbf{y}_i \approx \mathbf{t}_i$. If the target space is continuous then this is called regression, and if it is categorical (and therefore discrete) then this is called classification. The data is usually split in three portions; the training set, the cross-validation set and the test set. The training set is used to teach the machine about the correlations between the inputs and the targets. The test set is used to test the performance of the machine on data points which were not seen by the computer during the training phase. The cross-validation set is similar to the test set, but it should be used to tuned the hyper-parameters of the machine on data points not seen during the training phase. A classic example of supervised learning is hand-written digits recognition, where the data is images of hand-written digits labeled by the digit written on the image.

Unsupervised learning is about finding structure in unlabeled data. This usually means classifying our input data points $\mathbf{x}_i$ into a discrete set of clusters or applying some dimensionality reduction technique to our data points. This can be done for its own purpose, like if we are trying to categorize newspaper articles into some unknown categories for example, or it can be used in the context of supervised learning at the level of feature engineering. Indeed, if our data is noisy we can cluster our data points and then use only the center of the cluster as a feature in the supervised learning step.

Our goal in this section is to introduce a selection of topics on machine learning to the uninitiated. A more complete treatment can be found in [21, 22]. This section is organized as follows. We start by introducing some simple illustrative machine learning algorithm. Then we very briefly discuss what is arguably the most powerful supervised learning algorithm nowadays, namely neural networks. We end with a discussion of the previous applications of machine learning to particle physics. Later in this thesis we will build up on these applications by using a neural network in the context of top tagging.

1.2.1 Simple algorithms

To get an idea of what machine learning is about, we start by introducing one simple algorithm for supervised learning and one simple algorithm for unsupervised learning. These algorithms will help demonstrate the power of machine learning and it will be easy to generalize them to more complex algorithms which have particle physics applications.

Logistic regression

Let us start by introducing one of the simplest supervised learning algorithm, namely logistic regression. At its core logistic regression is simply the well known linear regression algorithm, also known as curve fitting, but for classification problems. With classification problems we are not trying to predict a continuous quantity, but instead we are trying to predict which category the input belongs to.

For concreteness let us work with a practical example. Suppose our data is a set of gray-scale images. On these images is either a cat or a dog. Our goal is to build a classifier which takes as input the image and produces a prediction for which animal is on the image. Our input is a vectorized version of an image $\mathbf{x}$, where each component corresponds

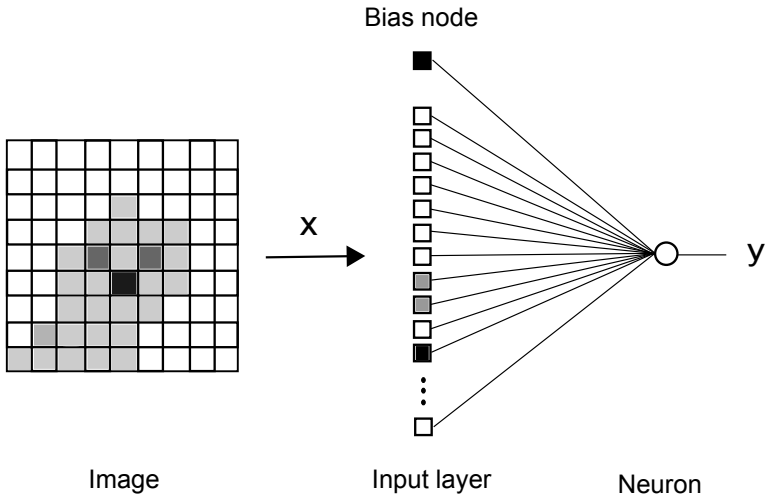

Figure 1.5: Diagram representing the logistic regression algorithm for the image classification problem of cats and dogs.

to the intensity of a pixel on the image. The target can then take the form of a single component vector, which takes the value of 0 for cats and 1 for dogs. If we decided to use linear regression for this problem our outputs would then take the form

$$y_i = \mathbf{w}^T \cdot \mathbf{x}_i + b \tag{1.23}$$

where the components of $\mathbf{w}$ are called the weights and b is a bias term. The problem with this approach is that the output may not lie in the target space $\in [0, 1]$. It is therefore not clear what a value of say $y_i = -2$ would mean. We can instead use logistic regression which uses a non-linear mapping to make sure the output is in the target space

$$y_i = \frac{1}{e^{-(\mathbf{w}^T \cdot \mathbf{x}_i + b)} + 1}. \tag{1.24}$$

This mapping gives a probabilistic interpretation for the output y. That is, we can think of y as the probability of $\mathbf{x}_i$ to be a dog image. In practice we could then classify all images which have $y > 0.5$ as images of dogs and all other images as images of cats. The architecture of the algorithm is illustrated on Fig. (1.5).

To quantify the quality of our classifier we need to define some error $E(w)$. This is a

useful step not only to assess the quality of our classifier but also because by using this error we can cast our machine learning problem as an optimization problem. Indeed, the algorithm which learns how to classify images of cats and dogs will simply be minimizing the error of the training data set. There are various options possible for the choice of error, but for logistic regression the most popular one is the log-loss

$$\text{Log-loss}(w) = -\frac{1}{N}\sum_{j=1}^{N}\left(t_i \log(y_i) + (1-t_i)\log(1-y_i)\right). \tag{1.25}$$

This choice of error is popular because it makes the optimization problem convex, meaning that there is only one minimum of the error in the weight space. Our goal is then to find the weights $\mathbf{w}$ which minimize the log-loss error. There are various optimization algorithms available, for simplicity we choose gradient descent. The idea behind gradient descent is to start at a random point in weight space, and then move in the direction of steepest descent by a small step. If we keep doing that iteratively we will eventually reach the minimum. Applying this to logistic regression gives

$$\mathbf{w}^{(n+1)} = \mathbf{w}^{(n)} - \eta\frac{\partial \text{Log-loss}}{\partial \mathbf{w}} = \mathbf{w}^{(n)} + \frac{\eta}{N}\sum_{j=1}^{N}(t_j - y_j)\,\mathbf{x}_j \tag{1.26}$$

$$b^{(n+1)} = b^{(n)} - \eta\frac{\partial \text{Log-loss}}{\partial b} = b^{(n)} + \frac{\eta}{N}\sum_{j=1}^{N}(t_j - y_j) \tag{1.27}$$

where η is a small learning rate.

One may argue that plain linear logistic regression is not powerful enough to distinguish cats from dogs. Indeed, at this point we only have a linear model, there is no non-linearity or interactions between the pixels. One way to fix this is to introduce quadratic interactions manually and add more weights accordingly,

$$y_i = \frac{1}{e^{-(\mathbf{w}^T\cdot\mathbf{x}_i + \sum_{jk} w_{jk}^{(int)}\mathbf{x}_i^{(j)}\mathbf{x}_i^{(k)} + b)} + 1} \tag{1.28}$$

where $\mathbf{x}_i^{(j)}$ is the jth component of the $\mathbf{x}_i$ vector. We could keep going like this and add cubic, quartic and other higher order interaction terms. However there is a danger to this

approach, namely over-fitting. Indeed, by adding more complexity to our model we risk fitting our model to random fluctuations as opposed to the underlying trend. When a classifier is overfitting the training error is very small but this will not generalize well to data sets which were not part of the training set. There are many known ways to avoid over-fitting with logistic regression, one of them is to use a regularization. The idea behind regularization is to add a penalty to the error which tries to keep the weights small. The new error is then

$$E(w) = \text{Log-loss}(w) + \frac{\lambda}{2}\left(\mathbf{w}^T \cdot \mathbf{w} + \sum_{jk}\left[w_{jk}^{(int)}\right]^2\right) \tag{1.29}$$

where λ is a parameter which needs to be fixed in a cross-validation phase.

Using all these pieces, we can summarize the logistic regression algorithm as follows:

1. Initialize the weights to small random values.
2. Repeat until convergence of log-loss:
   - Compute the output for each input vector using

$$y_i = \frac{1}{e^{-(\mathbf{w}^T \cdot \mathbf{x}_i + \sum_{jk} w_{jk}^{(int)} \mathbf{x}_i^{(j)} \mathbf{x}_i^{(k)} + \ldots + b)} + 1} \tag{1.30}$$

   - Adjust the weights at each iteration the following way:

$$\mathbf{w}^{(n+1)} = \mathbf{w}^{(n)}\left(1 - \eta \cdot \lambda\right) + \frac{\eta}{N}\sum_{i=1}^{N}\left(t_i - y_i\right)\mathbf{x}_i \tag{1.31}$$

$$w_{jk}^{(n+1)(int)} = \mathbf{w}_{jk}^{(n)(int)}\left(1 - \eta \cdot \lambda\right) + \frac{\eta}{N}\sum_{i=1}^{N}\left(t_i - y_i\right)\mathbf{x}_i^{(j)}\mathbf{x}_i^{(k)} \tag{1.32}$$

$$\ldots$$

$$b^{(n+1)} = b^{(n)} + \frac{\eta}{N}\sum_{i=1}^{N}\left(t_i - y_i\right) \tag{1.33}$$

K-means

Let us now discuss what is arguably the simplest unsupervised learning algorithm, namely K-means. Our goal with this algorithm is to arrange our data points into K clusters. As a concrete example, suppose we have some pictures in our data set which contains more than one animal, say some pictures have either 2 dogs or 2 cats. We do not want to pollute the training data for the supervised learning phase, but at the same time we do not want to throw away potentially useful data. A possible solution to this problem is to apply K-means on these images and then split the images using the cluster division.

Note that with K-means we must know the number of clusters, which is not always trivial. In our example, we know that the contaminated pictures contain either 2 dogs or 2 cats, so the number of clusters is $K = 2$. Once this is set, all that is left to do is to decide which points belong to which cluster, or in our case which pixels belongs to which animal. This decision will involve a distance measure, which we often choose to be Euclidean $d(\mathbf{a}, \mathbf{b}) = \sum_j (a_j - b_j)^2$ but depending on the problem we could choose something else with a small modification to the algorithm. Just like with logistic regression, K-means can be casted as an optimization problem. Indeed, K-means attempts to minimize the within-cluster sum of squares (WCSS), that is the sum of distances between the cluster centers μ^i and the points in their cluster:

$$WCSS = \sum_{i=1}^{K} \sum_{x \in S_i} \sum_{j} (x_j - \mu_j^i)^2 \tag{1.34}$$

where S_i are the cluster sets. Taking the first order partial derivative of the WCSS and setting to zero gives

$$\frac{\partial WCSS}{\partial \mu_j^i} = 2 \sum_{x \in S_i} (x_j - \mu_j^i) = 0 \quad \Rightarrow \quad \mu_j^i = \frac{1}{N_i} \sum_{x \in S_i} x_j \tag{1.35}$$

where N_i is the number of data points in the cluster S_i. This result is the basis of the K-means algorithm.

To use the K-means algorithm we start by positioning K centers randomly in our feature space. We can then assign each data point to their cluster by determining which centers they are the closest to. Based on the cluster assignments we can then recalculate the cluster centers according to Eq. (1.35). We repeat this procedure iteratively until the centers stop moving. This is summarized as follow:

1. Choose K.
2. Initilize K centers μ^i in the feature space.
3. Repeat until the centers μ^i stop moving:
    - Compute the distance $||\mathbf{x} - \mu^i||^2$ to each centers for each data point.
    - Assign each data point to their closest center $\min_i ||\mathbf{x} - \mu^i||^2$.
    - Recenter the cluster centers using

$$\mu_i = \frac{1}{N_i} \sum_{\mathbf{x} \in S_i} \mathbf{x} \tag{1.36}$$

In practice the K-means algorithm can get stuck at a local minimum of the WCSS. To solve this problem one usually runs the algorithm many times and pick the centers which minimize the WCSS.

In our cats and dogs example, we would assign a weight to each pixel equal to the pixel's intensity. Once we found the cluster centers using "2-means", we can then loop over the pixels and find their nearest cluster center to know which pixel describes which animal.

1.2.2 Neural networks

A neural network (NN) is a supervised learning algorithm just like logistic regression. However, while inserting complex non-linear interactions between the features is a com-

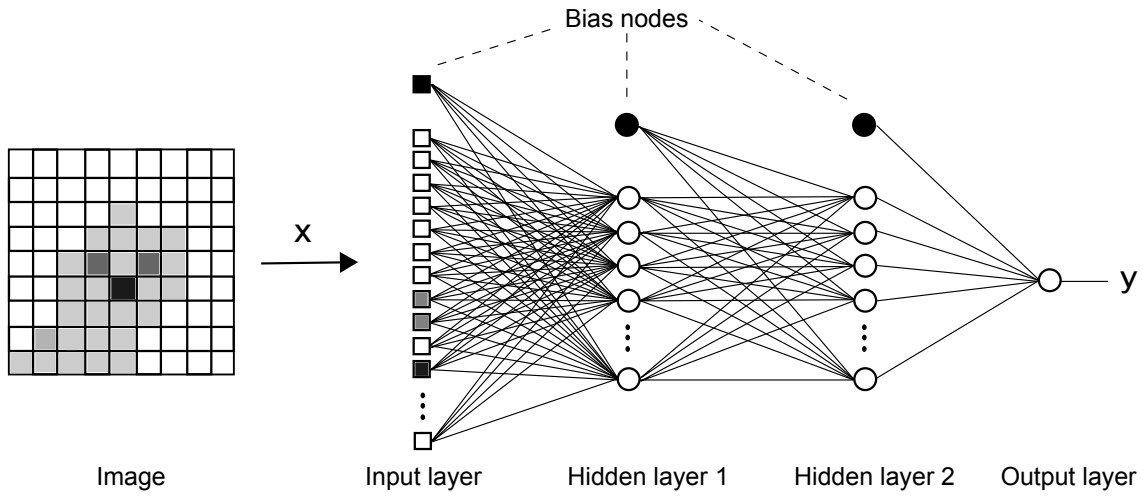

Figure 1.6: Graphical representation of an Artificial Neural Network (ANN).

plicated process in logistic regression, this is done naturally with NNs. NNs are therefore very useful when the underlying trend we are trying to model is intrinsically non-linear. The original motivation behind NNs was a biological one. Indeed, Frank Rosenblatt came up with the idea in 1958 as a way to emulate a biological brain [23]. In 1975 Paul Werbos found a way to train NNs which is called backpropagation [24] and this discovery made NNs one of the most popular learning algorithm ever since. In recent years, the field of NNs has seen a lot of progress since researchers found ingenious ways to train NNs with many layers. NNs are now used in many applications such as image recognition, sound recognition, automated driving, etc. To keep things practical in this section, we will keep using the cats and dogs image recognition example we used previously.

A simple architecture of a NN is shown in Fig. 1.6. If we think of logistic regression as a single neuron, then a NN is a set of layers with multiple neurons on each layers. The neurons are connected with the other neurons on the previous and later layers. Depending on the problem at hand we can adjust the architecture accordingly. Mathematically, the

NN can be thought of as a succession of non-linear transformations:[1]

$$\mathbf{x} \to h_i^{(1)} = f(W_{ij}^{(1)} x_j + b_i^{(1)}) \to \cdots \to h_i^{(l)} = f(W_{ij}^{(l)} h_j^{(l-1)} + b_i^{(l)}) \to y = f(W_j^{(O)} h_j^{(l)} + b^{(O)}), \tag{1.37}$$

where f is the so-called activation function, chosen to be the same as with logistic regression

$$f(z) = \frac{1}{1 + e^{-z}}. \tag{1.38}$$

The weights $W_{ij}^{(L)}$ and the biases $b_i^{(L)}$ are numbers determined by the training procedure.

Just like with logistic regression, training consists of adjusting the weights so that the actual outputs of the NN y_i is as close as possible to the target outputs t_i, across the training set. To quantify the error, we use the logarithmic loss function. We use the back-propagation algorithm, combined with gradient-descent minimization. The idea is very close to the gradient-descent algorithm used for logistic regression, as in we are still updating the weights using the same gradient descent rule. The difference here is that the derivative of the error function with respect to the weights can get more complicated for the earlier layers of the NN. However, by using partial derivatives we can simplify the calculation in such a way that the error at the output looks like it is backpropagated to the earlier layers. In its simplest version for a NN with 2 hidden layers, the algorithm can be summarized as follows [21]:

1. Initialize the weights of each neurons to small random values.
2. Repeat until convergence of log-loss, for each input vector $\mathbf{x}_i$:
    - Forward: Compute the output of each neuron until the output layer is reached, that is

$$\mathbf{x}_i \to h_i^{(1)} = f(W_{ij}^{(1)} \epsilon_j + b^{(1)}) \to h_i^{(2)} = f(W_{ij}^{(2)} h_j^{(1)} + b^{(2)}) \to y_i = f(W_j^{(O)} h_j^{(2)}) \tag{1.39}$$

[1] In Eq. (1.37) and below, repeated indices are always summed over.

- Backward: Adjust the weights of each neuron by propagating backward the error at the output using

$$\delta^{(O)} = (t - y_i)y_i(1 - y_i) \quad \text{and} \quad \delta_{\mathrm{k}}^{(\mathrm{l})} = \mathrm{h}_{\mathrm{k}}^{(\mathrm{l})}(1 - \mathrm{h}_{\mathrm{k}}^{(\mathrm{l})}) \sum_{\mathrm{j}} \mathrm{W}_{\mathrm{kj}}^{(\mathrm{l}-1)} \delta_{\mathrm{j}}^{(\mathrm{l}-1)} \tag{1.40}$$

$$\begin{aligned} W_k^{(0)} &\rightarrow W_k^{(0)} + \eta\delta^{(O)} h_k^{(2)} \\ W_{jk}^{(2)} &\rightarrow W_{jk}^{(2)} + \eta\delta_j^{(2)} h_k^{(1)} \\ W_{jk}^{(1)} &\rightarrow W_{jk}^{(1)} + \eta\delta_j^{(1)} \varepsilon_k \end{aligned} \tag{1.41}$$

Note that this version of the algorithm is using stochastic gradient descent as opposed to batch gradient descent. Indeed, with stochastic gradient descent one updates the weights after computing the error for a single data sample. This will eventually minimize the logloss error but in a stochastic way. We can easily make this algorithm deterministic by using batch gradient descent, that is we update the weights only after all the training data has been processed. Note also that while with logistic regression the log-loss error gives a convex optimization problem, meaning that there is a unique minimum, with NNs this is not the case. We can thus get stuck at a local minimum during the gradient descent training phase. Therefore, to reduce the odds of getting stuck at local minimum we can add what is known as a "momentum" to the updates. This means that the weights at iteration t, W_{ij}^t, are still being pushed by the update from the previous iteration ΔW_{ij}^{t-1}, for example

$$W_{ij}^t \rightarrow W_{ij}^t + \eta\delta_i^{(l)} h_j^{(l-1)} + \alpha\Delta W_{ij}^{t-1} \tag{1.42}$$

where $\alpha \in (0, 1)$ is a fixed parameter.

A major concern with NN classifiers is over-fitting the network to the training data. Just like with logistic regression, we can use a regularization on the weights of the NN. However in recent years several more sophisticated methods have been developed to reduce overfitting. One of them is known as dropout [25]. The idea behind dropout is to

drop randomly $\sim 50\%$ of the neurons at each iterations. Doing so ensures that the weights learned in each neuron do not depend heavily on the weights of the other neurons. Therefore each neuron learns something useful about the training data. It has even been proved that using dropout for a single layer NN is equivalent to averaging a large ensemble of NNs. Another method developed recently to avoid overfitting is to use unsupervised pretraining [26]. The idea here is to sequentially pretrain each layer of the NN using an unsupervised learning technique. For example, one can pretrain each layer with denoising auto-encoders. In that case, each layer learns to match a noisy version of the training data to the training data itself. Doing so insures that the neurons are learning something useful about the training data, that is how to reconstruct a noisy version of the data. After the pretraining phase is done we apply a regular backpropagation phase which should only fine tune the NN. Note that this method is also useful to avoid getting stuck at a local minimum and is therefore widely used for NNs with many layers.

1.2.3 Particle physics applications

In this section we present a small sample of the machine learning applications to particle physics. One of the classic application of machine learning to particle physics is for tagging algorithms. We therefore start by discussing two tagging algorithms which make use of interesting machine learning. We then discuss an online machine learning competition which involved particle physics.

Tagging algorithms

When the LHC produces heavy particles with large momentum, these particles' decay products form highly boosted jets. In order to probe new physics, it is important for experimentalists to be able to distinguish jets produced by heavy particles such as the

top quarks or W/Z bosons from the QCD initiated quark and gluon jets. This is called jet tagging. Obviously the jet mass is one observable used to distinguish the two kinds of jets. However, recently more sophisticated approaches which use the jet substructure information have been developed to improve the classification of jets. In this section we present two methods which are inspired by machine learning. The first method called N-subjetiness uses a variation of the K-means algorithm, while the second method uses Fisher discriminant which is similar to logistic regression.

N-subjetiness came about from the observation that the radiation pattern from say a top decay is very "3-prong" like compared to QCD initiated jets. N-subjetiness thus measures how well a jet can be decomposed into N clusters using essentially the K-means algorithm [27]. Thus, if we get a small 3 clusters decomposition error then we can tag the jet as a top jet. We define N-subjetiness $\widetilde{\tau}_N^{(\beta)}$ as a modification of the WCSS,

$$\widetilde{\tau}_N^{(\beta)} = \frac{1}{d_0}\sum_i p_{T,i}\min\left\{(\Delta R_{1i})^\beta,(\Delta \mathrm{R_{2i}})^\beta,...,(\Delta \mathrm{R_{Ni}})^\beta\right\} \tag{1.43}$$

$$d_0 = \sum_i p_{Ti}(R_0)^\beta \tag{1.44}$$

$$\Delta R_{J,i} = \sqrt{\Delta y_{J,i}^2 + \Delta\phi_{J,i}^2}. \tag{1.45}$$

Note that d_0 is simply a normalization factor where R_0 is the characteristic jet radius. $\Delta R_{J,i}$ is the distance in the rapidity-azimuth plane between a candidate cluster center (subjet) J and a constituent particle i. There are two differences with WCSS; first each data point is weighted by the transverse momentum of the particle $p_{T,i}$, second the distance measure between a data point and a cluster center is not necessarily squared. Indeed, only when $\beta = 2$ is the distance square. This β parameter needs to be tuned in a cross-validation phase. Note that once the optimal cluster centers are found, $\widetilde{\tau}_N$ describes how N-subjetty a particular jet is, such that $\frac{\widetilde{\tau}_3}{\widetilde{\tau}_2}$ is then a very good discriminant for top and QCD jets.

To find the optimal cluster centers, we modify the K-means algorithm to take into account the modifications we made to the WCSS. This results in a simple modification of

the update rule for the cluster centers:

$$y_J^{(n+1)} = \frac{\sum_{i \in S_i} p_{T,i} y_i \left[\left(y_i - y_J^{(n)} \right)^2 + \left(\phi_i - \phi_J^{(n)} \right)^2 \right]^{\frac{\beta-2}{2}}}{\sum_{i \in S_i} p_{T,i} \left[\left(y_i - y_J^{(n)} \right)^2 + \left(\phi_i - \phi_J^{(n)} \right)^2 \right]^{\frac{\beta-2}{2}}} \tag{1.46}$$

and similarly for $\phi_J^{(n+1)}$

N-subjetiness is a clever utilization of an unsupervised machine learning technique. One may wonder however if it is possible to use a supervised machine learning technique to do jet tagging. A group from SLAC recently devised a method which does precisely this for W jet tagging [28]. For the training set, they generated a large amount of jets using Monte Carlo simulation. These jets were then preprocessed so that they simulate the data collected by a calorimeter. The calorimeter data can then be seen as an image so that jet tagging becomes analogous to our cats and dogs example. The supervised learning algorithm they used is called Fisher's Linear Discriminant (FLD). FLD is a linear method similar to logistic regression. Note that while the performances of FLD are similar to the performances of logistic regression, their training procedures are significantly different. Indeed, the idea behind FLD is to identify the plane $\mathbf{F}$ in the high dimensional feature space (the pixels of our images) which maximizes the separation between the QCD jets and the W jets while simultaneously minimizing the scatter within each jet class. One can prove that such a plane is found by taking

$$\mathbf{F} = (\Sigma_W + \Sigma_Q)^{-1} (\mu_W - \mu_Q) \tag{1.47}$$

where μ_W (μ_Q) is the mean in feature space for the W (QCD) jets, while Σ_W (Σ_Q) is the covariance matrix in feature space for W (QCD) jets. Once we found the optimal separating plane $\mathbf{F}$ we can use it for tagging by projecting the jet image $\mathbf{A}$ we want to tag onto $\mathbf{F}$, and use the result $\sum_i A_i F_i$ as a discriminant. While this method gives good performances, note that FLD is sensitive to outliers and can only find linear correlations.

Kaggle: Higgs boson machine learning challenge

Kaggle is a website that hosts machine learning competitions. The setup is almost always the same; anyone can join a competition, download the training data related to that competition, code up a learning algorithm on their computer, produce predictions on the separate test data and finally upload the predictions to Kaggle. Kaggle then gives the user a score based on the accuracy of the predictions and a leaderboard is created. There is also usually some prices awarded to the top 3 competitors. For the users, it is a good way to practice their machine learning skills, and for the companies hosting the competitions it is a good way to outsource some of their data science problems.

Atlas used the Kaggle platform from May to September of 2014 to host a competition [29]. The goal was simple, come up with a classifier that can distinguish the Higgs boson from background. This is a classic problem in particle physics where machine learning is used. The channel of interest for this contest was Higgs decaying into two taus, and in particular events where one tau decays into an electron or a muon and two neutrinos, and the other tau decays in hadrons and a neutrino. This means that if such a decay occurs the Atlas detector can then collect information about a lepton coming from a tau decay, another reconstructed hadronically decaying tau, missing energy and jets. Note that the Higgs decaying into two taus is particularly difficult to identify because the background coming mainly from Z boson decay into two taus is similar to the decay of the Higgs but much more frequent.

The training data consisted of 250 000 events and the test data of 550 000 events, all of which were Monte Carlo simulated. Of these events, about $\sim 2/3$ were background while $\sim 1/3$ were signal (that is, produced by the Higgs). With each events came 30 features describing data that could be collected from the Atlas detector. These features were functions of the 4-momenta of the objects collected by the Atlas detector. For example, the

number of jets, the p_T, ϕ and η coordinates of the lepton, tau, leading and subleading jets were some of the features present.

This competition was a great opportunity for the machine learning community to interact with the particle physics community. With 1785 participants, including myself, this was the largest competition Kaggle had ever hosted at that time. The algorithms used by Atlas for this task were tree-based algorithms (such as random forest or gradient boosting machine), but the winner of the competition improved by 15% the performance of Atlas's classifier by using an ensemble of NNs.

1.2.4 Summary

In this section we provided a very selective review of some of the many topics that fall under the umbrella of machine learning. We focused only on a handful of topics which currently have applications in particles physics, though many other types of learning or algorithms could one day find a particle physics application. We reviewed logistic regression and K-means, both of which are used in jet substructure tagging algorithms. We then discussed neural networks which are widely used in image recognition tasks. They are also often used in multivariate analysis problems like the problem of the Kaggle Higgs competition.

1.3 Outline

This thesis covers research conducted between 2010 and 2015 at Cornell University. The subsequent chapters are based on papers written with collaborators on the topic of dark matter and the application of machine learning to top tagging.

Self-interaction became a popular topic in the dark matter community after it was realized that Sommerfeld enhancement could be used to explain the excess observed by PAMELA and Fermi. The type of self-interaction studied was however restricted to Coulomb and Yukawa since they were both easy to handle numerically. In chapter 2 we extend the type self-interaction possible to include both singular and spin-dependent interactions. These interactions often come up in models of composite dark matter. By using tools first developed in nuclear physics we show how to compute the Sommerfeld enhancement with these difficult interactions.

In chapter 3 we construct a concrete composite dark matter model. The idea behind the model is that the dilaton is the only particle that interacts with both dark matter and the standard model. This gives an economical model which only has 3 free parameters, namely the conformal breaking scale f, the dilaton mass m_σ and the dark matter mass m_χ. It also has a natural self-interaction mediator with the dilaton. We explore the parameter space of the model and show that a wide range of parameters are allowed in spite of the various experimental constraints.

Finally in chapter 4 we switch from dark matter to machine learning. Previously, the only supervised learning technique used for tagging was the Fisher discriminant which is a linear technique. In order to use the non-linear correlations which may be present in the jets substructure we expend on this idea by considering artificial neural networks. We therefore teach neural networks how to differentiate QCD jets from boosted top jets using a large amount of simulated hadronic calorimeter images. We also show how this approach systematically outperforms other tagging algorithms.

CHAPTER 2

THE EFFECTIVE THEORY OF SELF-INTERACTING DARK MATTER

While dark matter does not interact much with ordinary matter, the dark sector could have additional forces that attract or repulse dark matter particles themselves. These forces can change the rate at which dark matter particles annihilate into ordinary matter, thereby modifying dark matter's experimental signature. In this chapter we study these self-interacting forces and show how to calculate the modification of the annihilation rate of dark matter.

This chapter is based on the paper *The effective theory of self-interacting dark matter* [30], written in collaboration with Brando Bellazzini and Philip Tanedo.

2.1 Overview

We present an effective non-relativistic theory of self-interacting dark matter. We classify the long range interactions and discuss how they can be generated from quantum field theories. Generic dark sectors can generate singular potentials. We show how to consistently renormalize such potentials and apply this to the calculation of the Sommerfeld enhancement of dark matter interactions. We explore further applications of this enhancement to astrophysical probes of dark matter including the core vs. cusp problem.

2.2 Introduction

Less than a quarter of the matter density of the universe is composed of ordinary baryons. The remaining component is called dark matter (DM) and has only been probed through its gravitational interactions at cosmological and astrophysical scales. One appealing class of DM candidates are weakly-interacting massive particles (WIMPs). These are

- stable or long-lived compared to the age of universe
- non-relativistic upon freeze out from thermal equilibrium in the early universe
- electrically neutral and weakly interacting, i.e. with annihilation cross sections in the pb range, so that $\Omega_{\rm DM} h^2 \approx 0.1 \text{ pb}/\langle \sigma v \rangle$.

These features hint at a possible link between the cosmological properties of DM and the mechanism for electroweak symmetry breaking.

In principle, WIMP annihilations still occur today in dense regions of our galaxy. This type of indirect detection has gained attention recently due to possible anomalies in the positron fluxes measured by PAMELA [16], FERMI [31] and AMS-02 [32], and the gamma

ray spectrum measured by FERMI [33, 15, 34, 35, 36, 37]. Such signals, however, require the total WIMP annihilation cross section to be well much larger than their thermal value. Nevertheless, there are mechanisms to boost the annihilation rate to the level of experimental sensitivity without spoiling the relic abundance. One possibility is that DM has long range self-interactions mediated by a light force carrier. If this exchange of particles produces an attractive self-interaction, it can effectively increase the annihilation cross section because of Sommerfeld enhancement or resonance scattering [38, 39, 40, 41, 18, 42, 43]. The annihilation cross section is thus enhanced by a boost factor, $S\sigma_0$, with $S \geq 1$, where σ_0 is the short-range annihilation cross section.

More recently, self-interacting DM has recently been proposed as a viable solution to possible discrepancies between observations of small scale structures and the predictions from N-body simulations based on collisionless cold DM [44, 45] . In particular, dwarf galaxies show flat DM density profiles in halo cores, whereas collisionless cold DM predicts cusp-like profiles [46, 47]. In addition to this "core vs. cusp problem," there are the "missing satellites problem" and the "too big to fail problem," see e.g. [17] and references therein. While it is possible that these problems could be addressed by including baryonic physics to collisionless DM simulations [48], self-interacting DM offers a viable and motivated alternative scenario that is rich of interesting observational consequences [46, 47, 49].

The standard approach to self-interactions and Sommerfeld enhancement is to assume an ultra-light elementary scalar or vector ϕ in the dark sector which mediates a force between the DM particles [50, 51, 17]. In this paper we take a more agnostic approach; we construct an effective theory that only assumes rotationally invariant self-interactions in the dark sector. One can parameterize the possible non-relativistic potentials in terms of the DM mass m_χ, spin $\mathbf{s}$, transferred momentum $\mathbf{q}$, and relative velocity $\mathbf{v}$. We work at the leading order in the exchanged momentum and velocity which is an excellent approxima-

tion for cold DM. For example, we show in Section 2.3.2 that the most general long-range P- and T- symmetric potential between two DM particles of arbitrary spin, is

$$V_{\rm eff}^{P,T} = \frac{1}{4\pi r}\left[\widetilde{g}_1(r) + \widetilde{g}_2(r)(\mathbf{s}_1\cdot\mathbf{s}_2) + \frac{\widetilde{g}_3(r)}{\Lambda^2 r^2}\left(3\mathbf{s}_1\cdot\hat{r}\ \ \mathbf{s_2}\cdot\hat{r} - \mathbf{s}_1\cdot\mathbf{s}_2\right) + \frac{\widetilde{g}_{7,8}(r)}{\Lambda r}(\mathbf{s}_1 \pm \mathbf{s}_2)(\hat{r}\times\mathbf{v})\right] \tag{2.1}$$

where $\widetilde{g}_i(r)$ are arbitrary functions that depend only on the the DM separation, and Λ is the characteristic interaction scale of the dark sector that we take much larger than the mediator mass. At scales where the mediator mass can be neglected and the theory is weakly coupled, the couplings freeze to constants, $\widetilde{g}_i(r) \to g_i$.

Strongly interacting mediators in the dark sector can generate singular potentials through non-standard propagators, see e.g. [52, 53]. Notice, however, that even weakly coupled models can generate potentials that are more singular than the $1/r^2$ centrifugal barrier at short distances. For example, dark matter interactions mediated by a light pseudo-scalar produce a $\widetilde{g}_3$ term in the potential (2.1) which goes like $1/r^3$. This can be generated, for example, by Goldstone bosons [54]. Another example is DM with dipole interactions generated by charged states. These produce a $\widetilde{g}_3$ term in the potential. Models based on these magnetic dipole interactions were recently proposed as a way to resolve discrepancies between tentative signals in direct detection experiments [55]. More exotic potentials can be generated by the loop-level exchange of composite operators made of light fields [56, 57, 58, 59]. Table 2.1 shows examples of weakly coupled models, preserving P and T, that generate the various g_i in (2.1).

Such singular potentials must be regularized at short distances and then renormalized by requiring that low-energy observables are cutoff independent. We carry out this renormalization program and show how to extract physical predictions from singular potentials generated by DM self-interactions. In particular, we calculate the Sommerfeld enhancement from a $1/r^3$ potential, extending the analysis in [60] by including wave-

Interaction	g_1	g_2	g_3	g_7	g_8
$\bar{\chi}\chi\varphi$	✓	X	X	✓	X
$\bar{\chi}\gamma^5\chi\varphi$	X	X	✓	X	X
$i\bar{\chi}\gamma^\mu\gamma^5\chi\partial_\mu\varphi$	X	X	✓	X	X
$\bar{\chi}\gamma^\mu\chi A_\mu$	✓	X	X	✓	X
$i\bar{\chi}\gamma^5\gamma^\mu\chi A_\mu$	X	X	✓	X	X
$i\bar{\chi}\sigma^{\mu\nu}\chi F_{\mu\nu}$	X	X	✓	X	X

Table 2.1: Leading order P- and T-preserving long-range static potentials in (2.1) from massless real scalar φ, vector gauge boson A_μ, or field strength $F_{\mu\nu} = \partial_{[\mu}A_{\nu]}$ mediators. Observe that g_2 is not generated in the massless limit. g_8 is not generated because of the spin conservation in CP-symmetric theories of spin-$\frac{1}{2}$ DM. See Table 2.2 and 2.3 for more details.

function renormalization[1]. We plot the elastic scattering cross section as a function of the velocity and the mass near the resonance region where the boost factor is large. Astrophysical systems at various scales, from clusters to dwarf galaxies with velocities ranging from $v \sim 10^{-3}$ and $v \sim 10^{-5}$, provide constraints on the DM self-interactions and hence the Sommerfeld enhancement [50, 51, 17]. We leave a detailed investigation of how these bounds may be adapted to singular potentials for future work, but we point out that the formalism presented here may be useful to avoid these constraints because of the velocity dependence of the elastic cross section.

Even though Sommerfeld enhancement is typically relevant only for s-wave annihilations due to the centrifugal barrier, the self-interacting DM potential (2.1) does not generically conserve orbital angular momentum $\mathbf{L}^2$. Interaction channels with different orbital angular momenta, ℓ, can be coupled. This explains why the $\widetilde{g}_3$ term in (2.1), which would average to zero because of the isotropy of $\ell = 0$ states, can still be relevant for Sommerfeld enhancement in $\Delta\ell = 2$ transitions [60]. Moreover, spin-spin interactions with $\widetilde{g}_3 \neq 0$ in (2.1) may generate macroscopic long range interactions when the DM spins are polarized

[1] We note that wavefunction renormalization is essential for Sommerfeld enhancement to be cutoff independent. The numerical results in Section 2.5 match [60] within an order of magnitude for a specific choice of renormalization conditions.

on average [59].

This paper is organized as follows. In section 2 we derive an effective long-range, non-relativistic potential for self-interacting dark matter at leading order in WIMP velocity. In section 3 we present a procedure to renormalize singular potentials and apply this to the calculation of the physical, cutoff-independent Sommerfeld enhancement. In sections 4 and 5 we present numerical results for a $1/r^3$ potential and discuss the types of astrophysical bounds that such an analysis may be applied to. We conclude in section 6 and include appendices reviewing the standard procedure for calculating Sommerfeld enhancement for non-singular potentials and a convenient square well approximation for singular potentials.

2.3 Effective long-range potential

The elastic scattering amplitude $\mathcal{M}$ from rotationally invariant DM self-interactions is a scalar function of the spins $\mathbf{s}_i$, exchanged momentum $\mathbf{q}$, and relative velocity $\mathbf{v}$. It is often convenient to use the Hermitian operators $i\mathbf{q}$ and the velocity transverse to the momentum transfer,

$$\mathbf{v}_\perp = \mathbf{v} - \frac{\mathbf{q}(\mathbf{q}\cdot\mathbf{v})}{\mathbf{q}^2} = \mathbf{v} + \mathbf{q}/m_\chi \tag{2.2}$$

where the last equality follows from four-momentum conservation.

In the center of mass frame, the elastic scattering amplitude is

$$\mathcal{M} = \frac{-1}{\mathbf{q}^2 + m_\phi^2} \sum_i g_i(\mathbf{q}^2/\Lambda^2, \mathbf{v}_\perp^2)\mathcal{O}_i(\mathbf{s}_j \cdot i\mathbf{q}/\Lambda, \mathbf{s}_j \cdot \mathbf{v}_\perp, \mathbf{s}_1 \cdot \mathbf{s}_2) \tag{2.3}$$

where Λ is the heavy scale of the dark sector, e.g. the DM mass m_χ, and $\mathcal{O}_i$ are the spin matrix elements. We explicitly pull out a factor associated with the propagator for the light

force carrier with mass $m_\phi^2 \ll \mathbf{q}^2 \ll \Lambda^2$ which acts as an infrared (IR) regulator at large distances. Further, we only consider the leading term in the exchanged momentum $\mathbf{q}/\Lambda$ and DM velocities, which we assume to be small $v\,, v_\perp \ll 1$. This is a good approximation for cold DM in the phenomenologically interesting regime from dwarf galaxy scales $v \sim 10^{-5}$ to freeze out $v \sim 0.3$. This type of non-relativistic effective theory was recently applied to the direct detection of dark matter in [61, 62]. In order to conserve DM energy (and the total angular momentum) we assume that mediator bremsstrahlung is kinematically suppressed, $m_\chi \mathbf{v}^2 \ll m_\phi$. In other words, we work in the regime

$$\mathbf{v}^4 \ll \frac{m_\phi^2}{m_\chi^2} \ll \frac{\mathbf{q}^2}{m_\chi^2} \sim \mathbf{v}^2\,. \tag{2.4}$$

We assume mediators with spin less than 2 since the longitudinal components of massive particles with higher spins spoil the derivative expansion at scales comparable with their mass, $\mathbf{q} \sim m_\phi$.

2.3.1 Rotationally invariant non-relativistic operators

Under parity and time reversal, velocities, spins, and momentum transform as

$$P:\ i\mathbf{q} \to -i\mathbf{q}\,, \qquad \mathbf{s} \to +\mathbf{s}\,, \qquad \mathbf{v}_\perp \to -\mathbf{v}_\perp\,, \tag{2.5}$$

$$T:\ i\mathbf{q} \to +i\mathbf{q}\,, \qquad \mathbf{s} \to -\mathbf{s}\,, \qquad \mathbf{v}_\perp \to -\mathbf{v}_\perp\,. \tag{2.6}$$

In turn, one can build the following invariant parity-even operators

$$\mathcal{O}_1 = 1 \tag{2.7}$$

$$\mathcal{O}_2 = \mathbf{s}_1 \cdot \mathbf{s}_2 \tag{2.8}$$

$$\mathcal{O}_3 = -\frac{1}{\Lambda^2}(\mathbf{s}_1 \cdot \mathbf{q})(\mathbf{s}_2 \cdot \mathbf{q}) \tag{2.9}$$

$$\mathcal{O}_4 = (\mathbf{s}_1 \cdot \mathbf{v}_\perp)(\mathbf{s}_2 \cdot \mathbf{v}_\perp) \tag{2.10}$$

$$\mathcal{O}_{5,6} = -\frac{i}{\Lambda}\left[(\mathbf{s}_1 \cdot \mathbf{q})(\mathbf{s}_2 \cdot \mathbf{v}_\perp) \pm (\mathbf{s}_1 \cdot \mathbf{v}_\perp)(\mathbf{s}_2 \cdot \mathbf{q})\right] \tag{2.11}$$

$$\mathcal{O}_{7,8} = -\frac{i}{\Lambda}\left[(\mathbf{s}_1 \pm \mathbf{s}_2) \cdot (\mathbf{q} \times \mathbf{v})\right] , \tag{2.12}$$

where spin wavefunctions are suppressed for simplicity. Operators $\mathcal{O}_{5,6}$ respect parity but break time reversal. In the following we discard $\mathcal{O}_4$ because it is only generated by spin-2 mediators [62]. Relaxing parity invariance introduces eight additional operators [59]: four of those respect time reversal or, equivalently, CP

$$\mathcal{O}_9 = -\frac{1}{\Lambda}(\mathbf{s}_1 \times \mathbf{s}_2) \cdot i\mathbf{q} \,, \tag{2.13}$$

$$\mathcal{O}_{10,11} = (\mathbf{s}_1 \pm \mathbf{s}_2) \cdot \mathbf{v}_\perp \,, \tag{2.14}$$

$$\mathcal{O}_{12} = -\frac{i}{\Lambda}[\mathbf{s}_1 \cdot (\mathbf{q} \times \mathbf{v})](\mathbf{s}_2 \cdot \mathbf{v}_\perp) + \frac{i}{\Lambda}[\mathbf{s}_2 \cdot (\mathbf{q} \times \mathbf{v})](\mathbf{s}_1 \cdot \mathbf{v}_\perp) \,, \tag{2.15}$$

while other four break both P and CP

$$\mathcal{O}_{13,14} = -\frac{1}{\Lambda}(\mathbf{s}_1 \pm \mathbf{s}_2) \cdot i\mathbf{q} \,, \tag{2.16}$$

$$\mathcal{O}_{15} = (\mathbf{s}_1 \times \mathbf{s}_2) \cdot \mathbf{v}_\perp \,, \tag{2.17}$$

$$\mathcal{O}_{16} = -\frac{1}{\Lambda^2}(\mathbf{s}_2 \cdot \mathbf{q})[\mathbf{s}_1 \cdot (\mathbf{q} \times \mathbf{v})] + \frac{1}{\Lambda^2}(\mathbf{s}_1 \cdot \mathbf{q})[\mathbf{s}_2 \cdot (\mathbf{q} \times \mathbf{v})] \,. \tag{2.18}$$

Observe that self-conjugate DM is symmetric under the exchange $1 \leftrightarrow 2$. This is equivalent to invariance under $(\mathbf{q}, \mathbf{v}, \mathbf{s}_1) \leftrightarrow (-\mathbf{q}, -\mathbf{v}, \mathbf{s}_2)$, which forbids $\mathcal{O}_{6,8,10,12,13,16}$.

2.3.2 The general effective potential

A more general approach is to replace the free propagator with a general two-point function in (2.3). This may include arbitrary negative powers of $\mathbf{q}^2$ from non-local interactions mediated by light states that have been integrated out. In an integral representation, the amplitude is

$$\mathcal{M} = -\int_0^\infty d\mu^2 \frac{\rho(\mu^2)}{\mathbf{q}^2+\mu^2} \sum_i g_i(\mathbf{q}^2/\Lambda^2, \mathbf{v}_\perp^2)\mathcal{O}_i(\mathbf{v}_j \cdot i\mathbf{q}/\Lambda, \mathbf{s}_i \cdot \mathbf{v}_\perp, \mathbf{s}_1 \cdot \mathbf{s}_2) \tag{2.19}$$

where $\rho(\mu^2)$ is the spectral density of the theory which provides a common language to describe weakly and strongly coupled models. The standard propagator is recovered when $\rho(\mu^2) = \delta(\mu^2 - m_\phi^2)$.

Since the couplings always appear with the mediator's propagator, we can make the replacement $g_i(\mathbf{q}^2/\Lambda^2, \mathbf{v}_\perp^2) = g_i(-\mu^2/\Lambda^2, \mathbf{v}_\perp^2)$ after neglecting short-range interactions such as δ-functions. Moreover, for light mediators, the spectral density only has support for $\mu^2 \ll m_\chi^2, \Lambda^2$ so that we may further write $g_i(\mathbf{q}^2/\Lambda^2, \mathbf{v}_\perp^2) \simeq g_i(0,0) \equiv g_i$ unless this order vanishes. In such a case one should go to the leading non-vanishing order, $g_i \to (-\mu^2/\Lambda^2)^n g_i^{(n)}/n!$. We have also dropped the velocity dependence because it does not provide the leading contribution unless one fine tunes the coefficients of the UV operators to cancel the velocity-independent contributions [61, 62].

Taking the Fourier transform of the scattering amplitude with respect to $\mathbf{q}$, one obtains the long-range effective potential as a function of the relative distance $\mathbf{r}$ and velocity $\mathbf{v}$. For example, P- and T-symmetric interactions result in an effective long-range potential

$$V_{\text{eff}}^{P,T} = \frac{1}{4\pi r}\left[\widetilde{g}_1(r) + \widetilde{g}_2(r)(\mathbf{s}_1 \cdot \mathbf{s}_2) + \frac{\widetilde{g}_3(r)}{\Lambda^2 r^2}(3\mathbf{s}_1 \cdot \hat{r}\ \ \mathbf{s_2} \cdot \hat{r} - \mathbf{s}_1 \cdot \mathbf{s}_2) + \frac{\widetilde{g}_{7,8}(r)}{\Lambda r}(\mathbf{s}_1 \pm \mathbf{s}_2)(\hat{r} \times \mathbf{v})\right] \tag{2.20}$$

where $\widetilde{g}_i(r)$ are integrals of the Yukawa factor over the spectral density

$$\widetilde{g}_1(r) = \int_0^\infty d\mu^2 \rho(\mu^2) e^{-\mu r} \left(g_1 - g_1^{(1)} \frac{\mu^2}{\Lambda^2} \right) \tag{2.21}$$

$$\widetilde{g}_2(r) = \int_0^\infty d\mu^2 \rho(\mu^2) e^{-\mu r} \left[g_2 + \left(\frac{g_3}{3} - g_2^{(1)} \right) \frac{\mu^2}{\Lambda^2} \right] \tag{2.22}$$

$$\widetilde{g}_3(r) = g_3 \int_0^\infty d\mu^2 \rho(\mu^2) e^{-\mu r} \left[1 + \mu r + \frac{1}{3}(\mu r)^2 \right] \tag{2.23}$$

$$\widetilde{g}_{7,8}(r) = g_{7,8} \int_0^\infty d\mu^2 \rho(\mu^2) e^{-\mu r} \left(1 + \mu r \right) \tag{2.24}$$

It is understood that working to leading non-vanishing order, $g_{3,7,8}$ and $g_{1,2}^{(1)}$ should always be dropped unless the $\mathcal{O}(\mathbf{q}^0)$ terms like $g_{1,2}$ vanish or are suppressed. For weakly coupled dark sectors, $\widetilde{g}_{1,2}$ are the usual exponential factors while $\widetilde{g}_{3,7,8}$ carry additional polynomial corrections in the mediator mass. In general, these functions have an arbitrary r dependence, as expected when the mediator is a composite operator. A simple example is a four-fermi operator between spin-$\frac{1}{2}$ DM particles χ and a massless neutrino-like species ν, that is $\mathcal{L} = \sqrt{\alpha}[\bar{\nu}\gamma_\mu(1-\gamma_5)\nu][\bar{\chi}\gamma_\mu(a-b\gamma_5)\chi]$. The mediator is a composite operator made of two light fermions. It generates a singular potential at the loop level [57], $\widetilde{g}_{i=1,2} \propto 1/r^4$ and $\widetilde{g}_3 \propto \Lambda^2/r^2$. Note that the spin structure of the potential is fixed by the quantum numbers of the light mediator.

Note that for spin-$\frac{1}{2}$ DM the particle–antiparticle potential must have $g_8 = 0$ since CP corresponds to a factor $(-)^{S+1}$ and thus implies the conservation of total spin $\mathbf{S}^2 = (\mathbf{s}_1 + \mathbf{s}_2)^2$ which can only take values 0 and 1. In this case, it is convenient to express the potential in the following form

$$V^{(s_i=1/2)} = \frac{1}{4\pi r} \left\{ \left(\widetilde{g}_1(r) - \frac{3}{4}\widetilde{g}_2(r) \right) + \frac{1}{2}\widetilde{g}_2(r)\mathbf{S}^2 + \frac{\widetilde{g}_3(r)}{2\Lambda^2 r^2} \left[3(\mathbf{S}\cdot\hat{r})^2 - \mathbf{S}^2 \right] + \frac{2\widetilde{g}_7(r)}{m_\chi \Lambda\, r^2} \mathbf{S}\cdot\mathbf{L} \right\}, \tag{2.25}$$

where $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is the orbital angular momentum and $\mathbf{p} = m_\chi \mathbf{v}/2$ is the conjugate momentum, $[\mathbf{r}^i, \mathbf{p}^j] = i\delta^{ij}$.

At large distances that are still smaller than the mediator Compton wavelength, $\Lambda^{-1} \ll$

mediator	interaction	$\frac{1}{r}$	$\frac{1}{r}(\mathbf{s}_1 \cdot \mathbf{s}_2)$	$\frac{1}{r^3}\left[3(\mathbf{s}_1 \cdot \hat{r})(\mathbf{s_2} \cdot \hat{r}) - \mathbf{s}_1 \cdot \mathbf{s}_2\right]$
scalar	$\lambda_s \bar{\chi}\chi\varphi$	$-\lambda_s^2$	0	0
pseudoscalar	$i\lambda_p \bar{\chi}\gamma^5\chi\varphi$	0	$\frac{\lambda_p^2 m_\varphi^2}{3m_\chi^2}$	$\frac{\lambda_p^2}{m_\chi^2}h(m_\varphi, r)$
Goldstone	$\frac{1}{f}\bar{\chi}\gamma^\mu\gamma^5\chi\partial_\mu\varphi$	0	$\frac{4m_\varphi^2}{3f^2}$	$\frac{4}{f^2}h(m_\varphi, r)$
vector	$\lambda_v \bar{\chi}\gamma^\mu\chi A_\mu$	$\pm\lambda_v^2\left(1 + \frac{m_A^2}{4m_\chi^2}\right)$	$\pm\frac{2\lambda_v^2 m_A^2}{3m_\chi^2}$	$\mp\frac{\lambda_v^2}{m_\chi^2}h(m_A, r)$
axial vector	$\lambda_a \bar{\chi}\gamma^5\gamma^\mu\chi A_\mu$	0	$-\frac{8\lambda_a^2}{3}\left(1 - \frac{m_A^2}{8m_\chi^2}\right)$	$\lambda_a^2\left(\frac{1}{m_\chi^2} + \frac{4}{m_A^2}\right)h(m_A, r)$
field strength	$\frac{i}{2\Lambda}\bar{\chi}\sigma^{\mu\nu}\chi F_{\mu\nu}$	0	$\mp\frac{2m_A^2}{3\Lambda^2}$	$\pm\frac{1}{\Lambda^2}h(m_A, r)$

Table 2.2: Parity-preserving particle–(anti-)particle (upper/lower sign) long-range, static potentials from scalar φ, gauge boson A_μ, and field strength $F_{\mu\nu} = \partial_{[\mu}A_{\nu]}$ mediators. Here $\sigma^{\mu\nu} = \frac{i}{4}[\gamma^\mu, \gamma^\nu]$ and h is defined in (2.28). Each term implicitly carries a Yukawa factor $e^{-m_\phi r}/4\pi$. Observe that the long-range $\mathbf{s}_1 \cdot \mathbf{s}_2$ is always suppressed by the mediator mass since $\lambda_a = m_A/f$.

$r \ll \mu^{-1}$, the functions $\widetilde{g}_i(r)$ become constants and the potential simplifies even further:

$$V_{\text{eff}}^{P,T} = \frac{1}{4\pi r}\left[g_1 + g_2(\mathbf{s}_1 \cdot \mathbf{s}_2) + \frac{g_3}{\Lambda^2 r^2}(3\mathbf{s}_1 \cdot \hat{r}\ \ \mathbf{s_2} \cdot \hat{r} - \mathbf{s}_1 \cdot \mathbf{s}_2) + \frac{g_{7,8}}{\Lambda r}(\mathbf{s}_1 \pm \mathbf{s}_2)(\hat{r} \times \mathbf{v})\right] . \quad (2.26)$$

This is the regime where Sommerfeld enhancement may be effective because the interaction is still long-range compared to the short distance annihilation processes that take place at $r \sim \Lambda^{-1}$.

The expressions for the potentials that break P but respect T are presented in Appendix A.

2.3.3 Weakly coupled examples

As an example, consider a dark sector with a weakly coupled light scalar or vector mediator ϕ with interactions $\lambda\mathcal{O}^{\text{QFT}}$ in Table 2.2. These generate a static potential

$$\sum_i \lambda_i \mathcal{O}_i^{\text{QFT}} \longrightarrow V_{\text{s}}^P = \left[g_1 + g_2(\mathbf{s}_1\cdot\mathbf{s}_2) + \frac{g_3}{\Lambda^2 r^2} h(m_\phi, r)\left[3\left(\mathbf{s}_1\cdot\hat{r}\right)\left(\mathbf{s_2}\cdot\hat{r}\right) - \mathbf{s}_1\cdot\mathbf{s}_2\right]\right]\frac{e^{-m_\phi r}}{4\pi r}, \tag{2.27}$$

where h encodes the dependence on the mediator mass,

$$h(m_\phi, r) = \left(1 + m_\phi r + \frac{m_\phi^2 r^2}{3}\right). \tag{2.28}$$

Table 2.2 gives the contributions to each of the coefficients on the right-hand side of (2.27) coming from the corresponding types of QFT interactions.

Note that the Dirac DM mass m_χ breaks axial symmetry so that the limit of a massless axial gauge boson mediator is consistent at finite m_χ only when chiral symmetry is broken spontaneously at a scale f so that $m_A = \lambda_a f$. In this case the transverse components decouple, $\lambda_a = m_A/f \to 0$, and only the longitudinal modes contribute to the amplitude with coupling $1/f$, matching the result from Goldstone boson exchange.

Table 2.3 gives the long-range, non-static potential contributions from massless scalars and gauge bosons. The $\mathbf{v}_\perp^2$ contribution generates a $\sim 1/r(\mathbf{v}^2 + \hat{r}(\hat{r}\cdot\mathbf{v})\mathbf{v})$ in position space which can be neglected because it is always subleading. Pseudo-scalar, axial-vector and field strength mediators, give vanishing non-static, long-range potentials at this order. Note that these potentials generically need to be complemented by the relativistic corrections to the kinetic energies, $\mathbf{p}^2/m_\chi^2\left(1 - \mathbf{p}^2/(4m_\chi^2) + \ldots\right)$. In the following sections we neglect these corrections to the kinetic energy since we checked that their contribution is very small.

mediator	interaction	$\frac{1}{4\pi r}\left[\mathbf{v}^2+\hat{r}(\hat{r}\cdot\mathbf{v})\mathbf{v}\right]$	$\frac{1}{4\pi r^2}\left(\hat{r}\times\mathbf{v}\right)\cdot\left(\mathbf{s}_1+\mathbf{s}_2\right)$
scalar	$\lambda_s\bar{\chi}\chi\varphi$	$-\frac{\lambda_s^2}{8}$	$\frac{\lambda_s^2}{4m_\chi}$
vector	$\lambda_v\bar{\chi}\gamma^\mu\chi A_\mu$	$\pm\frac{\lambda_v^2}{8}$	$\mp\frac{3\lambda_s^2}{4m_\chi}$

Table 2.3: Parity-preserving particle–(anti-)particle (upper/lower sign) long-range, non-static potentials from massless scalars φ and gauge bosons A_μ. Long-range contributions from pseudo-scalars, axial vectors and field strength vanish for massless mediators.

2.4 Renormalization of singular potentials and Sommerfeld enhancement

The potential $V_{\text{eff}}^{P,T}$ in (2.20) represents the most general long-range interactions between DM particles that preserve parity and time reversal. A standard method for calculating the Sommerfeld enhancement for the non-singular Coulomb and Yukawa potentials is presented in [38] and reviewed in Appendix B. In practice, one determines the boost factor by solving a Schrödinger-like equation with the proper boundary conditions. However, since the terms in $V_{\text{eff}}^{P,T}$ are typically very singular, the usual calculations for the boost factor will generically fail. In this section we show how to overcome these problems by renormalizing the Schrödinger equation. Since a full numerical solution can be computationally intensive for singular potentials, we also provide an algebraic algorithm to estimate the Sommerfeld enhancement for general potentials in Appendix C.

2.4.1 Wilsonian treatment of divergences

Potentials that diverge faster than $1/r^2$ at the origin are called singular [63] and generically arise in dark sectors with spinning DM and/or with some strong dynamics. The occurrence of unphysical behavior originating from the infinitely large energies of such potentials are analogous to the infinities of quantum field theory (QFT). These inconsistencies arise when one extrapolates a long-range potential to arbitrarily short distances where ultraviolet physics should be taken into account. In fact, the Schrödinger equation can be renormalized by adopting the Wilsonian renormalization group (RG) methods of QFT [64]: the singular potential is regulated at a short distance a and augmented with a series of local operators that parametrize the unknown UV physics,

$$V(r) \longrightarrow V(r)\theta(r-a) + c_0(a)\delta^3(r) + c_2(a)a^2\nabla^2\delta^3(r) + \dots \tag{2.29}$$

The short-distance part of this effective potential is a derivative expansion that can be truncated to the desired order as long as the typical momenta q are much smaller than the cutoff scale $\Lambda = a^{-1}$. This given order in q determines the finite set of coupling constants $c_i(a)$ which can be determined by low-energy data.

2.4.2 Renormalized potential

Singular potentials diverge at the origin so that further care is required to impose boundary conditions. The Schrödinger equation for an ℓ-wave state is conveniently expressed using the dimensionless coordinate $x = pr$, the product of the dark matter relative momentum and separation:

$$-\Phi''_{p,\ell}(x) + \left(\mathcal{V}(x) + \frac{\ell(\ell+1)}{x^2} - 1\right)\Phi_{p,\ell}(x) = 0, \tag{2.30}$$

where the dimensionless potential is rescaled by the momentum p and reduced mass $M = m_\chi/2$,

$$\mathcal{V}(x) = \frac{2M}{p^2} V\left(\frac{x}{p}\right). \tag{2.31}$$

We regulate the potential at $x_{\text{cut}} = ap$ with a square well of height $\mathcal{V}_0$ encoding the UV data of the relativistic completion,

$$\mathcal{V}_{\text{reg}}(x) = \mathcal{V}(x)\,\theta(x - x_{\text{cut}}) + \frac{1}{x_{\text{cut}}^2}\mathcal{V}_0\,\theta(x_{\text{cut}} - x). \tag{2.32}$$

In practice, we simulate the local counter-terms with a short-distance square well potential which makes the calculations much easier [65]. We stress, however, that any other choice or deformation of the counter-terms is allowed and physically equivalent as long as it changes only the UV behavior of the interactions [64].

Observe that the centrifugal barrier is left uncut since it is non-singular and unrelated to the UV physics. Once $\mathcal{V}_0$ is known, one may integrate the Schrödinger equation subject to the usual boundary condition at zero

$$\lim_{x\to 0} \Phi_{p,\ell}(x) = x^{\ell+1}, \tag{2.33}$$

and then extract the Sommerfeld enhancement from the asymptotic solution. In the regulated region $x < x_{\text{cut}}$, the Schrödinger equation can be solved explicitly in the approximation $x_{\text{cut}} \ll 1$,

$$\Phi_{p,\ell}(x < x_{\text{cut}}) = \Gamma\left(\ell + \frac{3}{2}\right)\left(\frac{2x_{\text{cut}}}{\mathcal{V}_0^{1/2}}\right)^{\ell+1/2} x^{1/2}\, J_{\ell+1/2}\left(\mathcal{V}_0^{1/2}\frac{x}{x_{\text{cut}}}\right). \tag{2.34}$$

The value $\mathcal{V}_0$ that appears in the Schrödinger equation is determined by requiring that a low energy observable is independent of the particular choice of the cutoff, x_{cut}. It is thus meaningful to define $\mathcal{V}_0(x_{\text{cut}})$ with respect to the value of a physical observable, which can be conveniently chosen to be the scattering phase δ_ℓ of the elastic dark matter scattering process that generates this enhancement.

For the region $x > x_{\text{cut}}$, recall that the general solution to the Schrödinger equation is a linear combination of two independent solutions,

$$\Phi_{p,\ell}(x > x_{\text{cut}}) = Af(x) + Bg(x). \tag{2.35}$$

Asymptotically far from the origin, these independent solutions are combinations of sines and cosines. The scattering phase is related to the shift in the argument when the asymptotic solution is written as a pure sine. Thus the δ_ℓ has a one-to-one relation to the ratio A/B. In this way A/B contains the UV data that can be measured in a low energy observable, the scattering phase shift.

In order to determine $\mathcal{V}_0(x_{\text{cut}})$ subject to a fixed scattering phase, we may match the logarithmic derivatives of the two piecewise solutions at x_{cut}. Comparing (2.34) with (2.35),

$$-\frac{\ell}{x_{\text{cut}}} + \frac{\mathcal{V}_0^{1/2}(x_{\text{cut}})}{x_{\text{cut}}} \frac{J_{\ell-1/2}\left(\mathcal{V}_0^{1/2}(x_{\text{cut}})\right)}{J_{\ell+1/2}\left(\mathcal{V}_0^{1/2}(x_{\text{cut}})\right)} \;=\; \frac{\frac{A}{B}f'(x_{\text{cut}}) + g'(x_{\text{cut}})}{\frac{A}{B}f(x_{\text{cut}}) + g(x_{\text{cut}})}. \tag{2.36}$$

Observe that matching the logarithmic derivative gives an expression that depends on A/B which is cutoff independent and directly related to our low-energy observable [65]. Once $\mathcal{V}_0(x_{\text{cut}})$ is determined, (2.32) is the correct non-singular low-energy potential for the problem with the given cutoff.

Due to the oscillatory nature of the Bessel function, there can be multiple solutions to the transcendental equation (2.36). These solutions are physically equivalent. To simplify our calculations we choose the first quadrant so that $\mathcal{V}_0(x_{\text{cut}})$ can take values in the range $(-\infty, \mathcal{V}_{\text{max}})$ where $\mathcal{V}_{\text{max}}$ is given by the first positive solution of

$$J_{\ell+1/2}\left(\mathcal{V}_{\text{max}}^{1/2}\right) \;=\; 0 \tag{2.37}$$

For $\ell = 0$, $\mathcal{V}_{\text{max}} = \pi^2$.

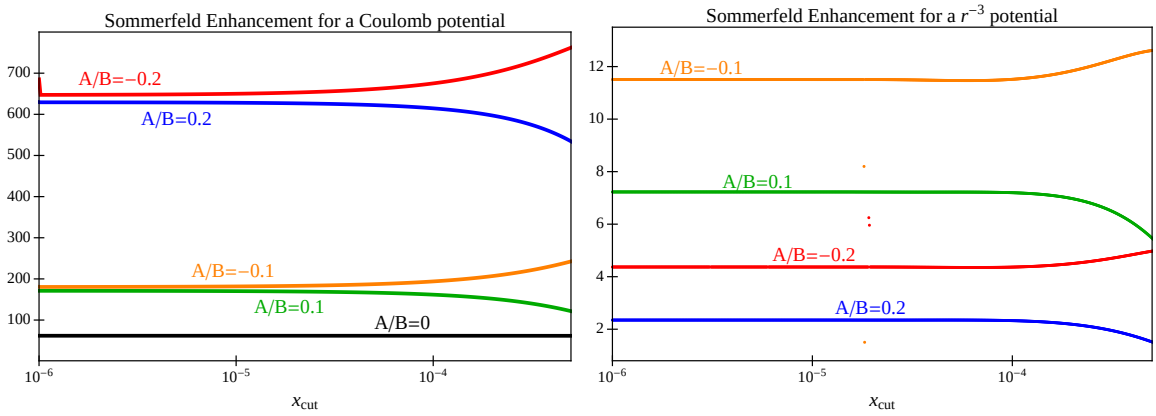

Figure 2.1: Cutoff-dependence of s-wave Sommerfeld enhancement using the procedure described in the text. Low energy data is encoded by the ratio A/B in (2.35). We take relative velocity $v = 10^{-3}$. Deviations from flatness reflect a breakdown of the $x_{\rm cut} \ll 1$ approximation. LEFT: Coulomb potential with $\alpha/v = e^2/4\pi v = 10$. The unique phase ($A/B = 0$) given by a QED-like UV completion is indicated by the black line. RIGHT: r^{-3} potential with $\widetilde{\alpha} = 2M^2 v\alpha/f^2 = 10^{-3}$, for α defined in (2.46).

2.4.3 Wavefunction renormalization

Since $\mathcal{V}_{\rm reg}$ in (2.32) is manifestly non-singular, one may proceed to solve the Schrödinger equation (2.30) subject to (2.33) following the procedure outlined in Appendix B. The resulting Sommerfeld enhancement, $S^{(0)}$, appears to depend on the choice of $x_{\rm cut}$. This residual cutoff dependence is not physical and is removed by including wavefunction renormalization, Z_ℓ:

$$S_\ell = Z_\ell S_\ell^{(0)} \tag{2.38}$$

Z_ℓ is fixed by using the observation that at relativistic speeds the Sommerfeld enhancement factor should go to one,

$$Z_\ell = \frac{1}{S_\ell^{(0)}(v \to 1)}. \tag{2.39}$$

2.4.4 Comparison to Coulomb potential

We now verify that the above procedure matches the usual result for the non-singular Coulomb potential, $V(r) = -\alpha/r$. The wavefunction in the region $x > x_{\text{cut}}$ is

$$\Phi_{p,\ell}(x > x_{\text{cut}}) = Ax^{1/2} J_{2\ell+1}\left(2\sqrt{\frac{x\alpha}{v}}\right) + Bx^{1/2} Y_{2\ell+1}\left(2\sqrt{\frac{x\alpha}{v}}\right). \tag{2.40}$$

One can check that the Sommerfeld enhancement is indeed independent of the choice x_{cut}. For different choices of A/B, one can obtain different Sommerfeld enhancements, as seen by the different lines on the left plot of Fig. 2.1. Of these, one line (black) corresponds to the analytical formulae found in the literature [38]; this corresponds to picking a scattering phase that is consistent with a relativistic completion that includes a massless boson. In other words, this is the choice that is consistent with a theory where the non-relativistic Coulomb potential is completed by a relativistic field theory resembling QED. Other choices correspond to theories whose non-relativistic limit is Coulomb but whose local interactions differ from pure QED.

2.5 Numerical results

The general DM potential considered here does not generally conserve orbital angular momentum $\mathbf{L}^2$ so that a coupled channel analysis between different ℓ-wave annihilation modes is required. This implies that the g_3 contribution in (2.1) can still be relevant for Sommerfeld enhancement via $\Delta\ell = 2$ transitions even though it averages to zero for $\ell = 0$ states [60]. This is contrary to the common belief that Sommerfeld enhancement is relevant only for s-wave annihilations due to the centrifugal barrier. For some states $\mathbf{L}^2$ is a well-defined quantum number once the total angular momentum J, the total spin S and parity $P = \pm$ are specified. In these cases the calculation of the boost factor reduces to a standard single-channel Schrödinger problem as discussed above. Table 2.4 shows the quantum

numbers for fermionic DM for low total angular momenta. Among the $\ell = 0$ states, ($J = 0$, $S = 0$, $P = -$) gives a single channel problem with arbitrary potential $V_0(r)$, whereas ($J = 1$, $S = 1$, $P = -$) requires a coupled channel analysis between $\ell = 0$ and $\ell = 2$.

J	S	P	ℓ
0	0	$-$	0
0	1	$+$	1
1	0	$+$	1
1	1	$+$	1
1	1	$-$	0, 2

Table 2.4: Low total angular momentum, J, DM scattering states labelled by spin, S, parity, P, and orbital angular momentum ℓ. J, S, and P are conserved by the Hamiltonian and are used to label states.

Assuming parity conservation, the effective potential $V_{\text{eff}} = \langle \text{out}|V(r)|\text{in}\rangle + \ell(\ell+1)/(2Mr^2)$ for each channel is obtained by sandwiching (2.25) with the centrifugal term between the appropriate $|\, J\ S\ P\,\rangle$ states,

$$|\,0\quad 0\quad -\rangle \;\to\; V_{\text{eff}} = \left(\widetilde{g}_1(r) - \frac{3}{4}\widetilde{g}_2(r)\right)\frac{1}{4\pi r} \tag{2.41}$$

$$|\,0\quad 1\quad +\rangle \;\to\; V_{\text{eff}} = \frac{1}{Mr^2} + \left(\widetilde{g}_1(r) + \frac{1}{4}\widetilde{g}_2(r) - \frac{\widetilde{g}_3(r)}{2\Lambda^2 r^2} - \frac{2\widetilde{g}_7(r)}{M\Lambda r^2}\right)\frac{1}{4\pi r} \tag{2.42}$$

$$|\,1\quad 0\quad +\rangle \;\to\; V_{\text{eff}} = \frac{1}{Mr^2} + \left(\widetilde{g}_1(r) - \frac{3}{4}\widetilde{g}_2(r)\right)\frac{1}{4\pi r} \tag{2.43}$$

$$|\,1\quad 1\quad +\rangle \;\to\; V_{\text{eff}} = \frac{1}{Mr^2} + \left(\widetilde{g}_1(r) + \frac{1}{4}\widetilde{g}_2(r) + \frac{\widetilde{g}_3(r)}{4\Lambda^2 r^2} - \frac{\widetilde{g}_7(r)}{M\Lambda r^2}\right)\frac{1}{4\pi r} \tag{2.44}$$

$$|\,1\quad 1\quad -\rangle \;\to\; V_{\text{eff}} = \frac{1}{Mr^2}\begin{pmatrix}0 & 0\\ 0 & 3\end{pmatrix} + \begin{pmatrix}\widetilde{g}_1(r) + \frac{\widetilde{g}_2(r)}{4} & \frac{\widetilde{g}_3(r)}{2\sqrt{2}\Lambda^2 r^2}\\ \frac{\widetilde{g}_3(r)}{2\sqrt{2}\Lambda^2 r^2} & \widetilde{g}_1(r) + \frac{\widetilde{g}_2(r)}{4} - \frac{\widetilde{g}_3(r)}{4\Lambda^2 r^2} - \frac{3\widetilde{g}_7(r)}{M\Lambda r^2}\end{pmatrix}\frac{1}{4\pi r} \tag{2.45}$$

where the $\ell = 0$ and $\ell = 2$ channels are coupled in (2.45). If the $\widetilde{g}_i$ are constant, then at leading order these channels are effectively non-singular and Coulomb-like. However, if $\widetilde{g}_1 + \widetilde{g}_2/4 = 0$, such as for pseudo-scalar exchange, then some of these channels are dominated by the singular $V \sim 1/r^3$ term. Moreover, one can also consider scenarios—for example, the exchange of multiple light particles [56, 57, 58, 59]—in which $\widetilde{g}_{1,2} \sim 1/r^3$ so that even the $\widetilde{g}_1$ and $\widetilde{g}_2$ terms are singular with $\ell = 0$. Thus one may in principle generate

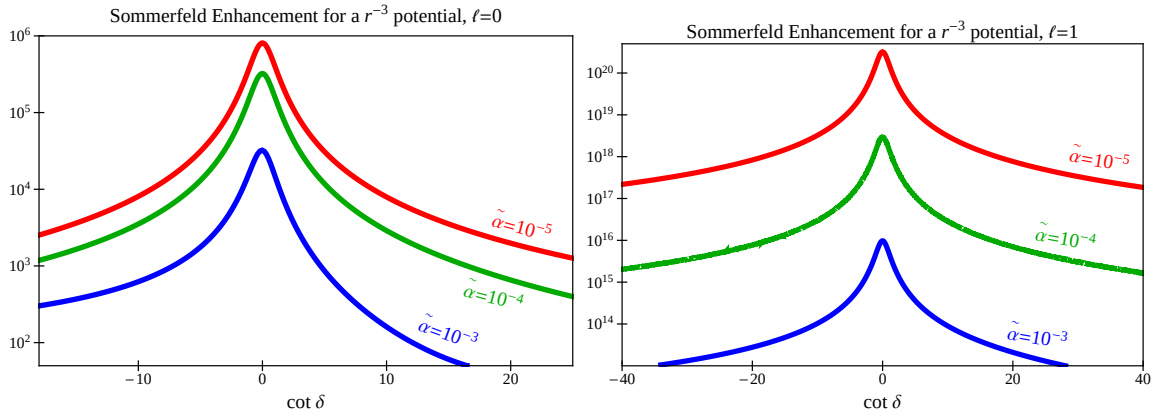

Figure 2.2: Sommerfeld enhancement for a singular r^{-3} potential and orbital angular momentum $\ell = 0$ (LEFT) and $\ell = 1$ (RIGHT) for relative velocity $v = 10^{-3}$ and various values of $\widetilde{\alpha} = 2M^2 v\alpha/f^2$, with α defined in (2.46).

a singular potential for any partial wave. For simplicity, we shall consider a simple $1/r^3$ potential for both $\ell = 0$ and $\ell = 1$. The coupled channel in (2.45), however, requires a more careful analysis that we leave for future work.

In Fig. 2.2 we plot the Sommerfeld enhancement for a potential

$$V(r) = -\frac{\alpha}{f^2 r^3} \tag{2.46}$$

as a function of the IR observable $\cot\delta$ for $\ell = 0, 1$. When comparing these, note that the $\ell = 1$ cross section has an additional factor of v^2 relative to $\ell = 0$. The resonance is located at $\cot\delta = 0$ because this is where the cross section is maximal. These plots can be used to give an upper bound on Sommerfeld enhancement for various couplings. Note that while it is true that the resonance is larger for smaller couplings, it requires more tuning from the UV to reach the resonance for a smaller coupling. Moreover, while $\cot\delta$ contains data about UV physics, it also depends on the IR coupling in such a way that reducing the coupling would not increase the Sommerfeld unless one simultaneously increases the height of the square well potential V_0.

Fig. 2.3 presents an exploration of these resonances as a function of the dark matter re-

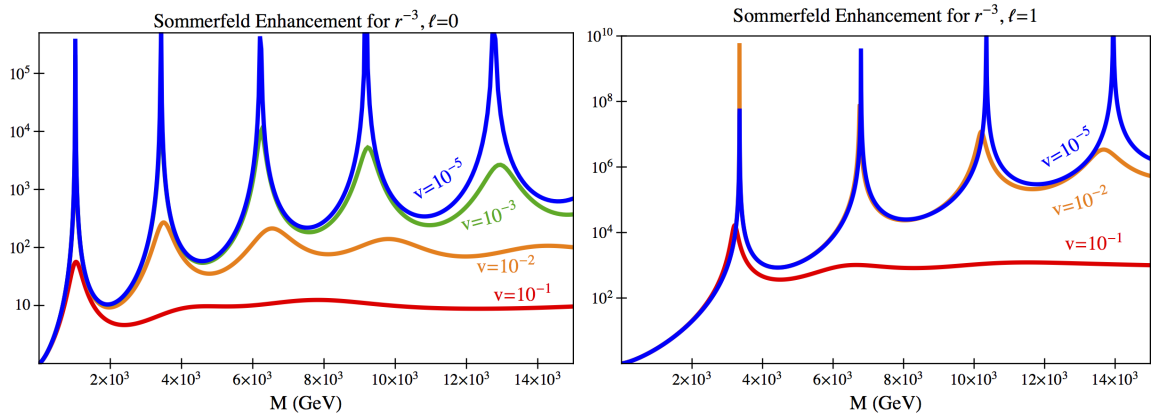

Figure 2.3: Resonances in Sommerfeld enhancement for a singular r^{-3} potential and orbital angular momentum $\ell = 0$ (LEFT) and $\ell = 1$ (RIGHT) for a range of relative velocities and α/f^2 = TeV^{-2} with α defined in (2.46). The large enhancements can be understood from the box approximation, see the Appendix C. For simplicity the height of the regulated potential is fixed by continuity with the long range piece.

duced mass. As described in the procedure above, the physical Sommerfeld enhancement for a singular potential requires information from an IR observable such as the scattering phase δ. As a reasonable estimate for natural UV models, we regulate the theory at a length scale r_0 where the non-relativistic description breaks down, $V(r_0) = M$. We then fix the height of the cutoff by continuity with the singular long-range part, $V_0 = V(r_0) = M$. Notice that for a $V(r) = -\alpha/(f^2 r^3)$ potential with $f = 1$ TeV, the dark matter mass necessary to reach a significant enhancement is about 1 TeV. If the dark matter mass is sufficiently large one may also need to consider the $\ell = 1$ contribution. This appears to contradict the common belief that $\ell > 0$ enhancement is too velocity suppressed to be relevant.

2.6 Phenomenology

While the collisionless cold DM paradigm successfully accounts for the large scale structure of the universe, it faces tension at smaller scales where N-body simulations present some discrepancies with observations. In particular, dwarf galaxies show flat core DM densities profiles in the central part of the halos, whereas collisionless cold DM predicts cusp-like profiles [66, 67, 68, 69]. While this discrepancy may be due to unaccounted baryonic physics [70, 71, 72], it may alternately be taken as a motivation for dark matter self-interactions [73, 47, 46]. A related astrophysical motivation for self interactions is the "too big to fail problem," in which the brightest observed dwarf spheroidal satellites in the Milky Way appear to be incompatible with the central densities of subhalos predicted by collisionless DM[74, 75, 76]. A third suggestion for self interactions is the "missing satellites problem"; collisionless DM predictions for the number the satellite galaxies expected in the Milky Way appears to disagree with observations [77, 78]. See, e.g. [50, 17] and references therein for critical discussions or [79] for an alternate solution that invokes DM.

To solve the core vs. cusp problem, the dark matter self interaction must have a sufficiently large cross section, $\sigma/m_\chi \sim 0.1 - 10$ cm^2/g, for velocities typical of dwarf galaxies, $v \sim 10^{-5}$, while having a smaller cross section for galaxy cluster velocities, $v \sim 10^{-3}$, where collisionless DM results are in good agreement. There are additional upper bounds on the cross section coming from astrophysical observations sensitive to the velocities characteristic of galaxy clusters [50, 51]. One of the most stringent bounds, for example, comes from the ellipticity of galaxy clusters [80, 51, 81]. The most recent simulations have softened this bound to $\sigma/m_\chi = 0.1$ cm^2/g [46, 47]. Further, the cosmic microwave background (CMB) sets an upper bound on Sommerfeld enhancement from the effect of DM annihilation after recombination [82, 83, 84]. Though a constant cross section $\sigma/m_\chi \lesssim 0.5$ cm^2/g may account for these effects, this velocity dependence is also suggestive of a Sommerfeld

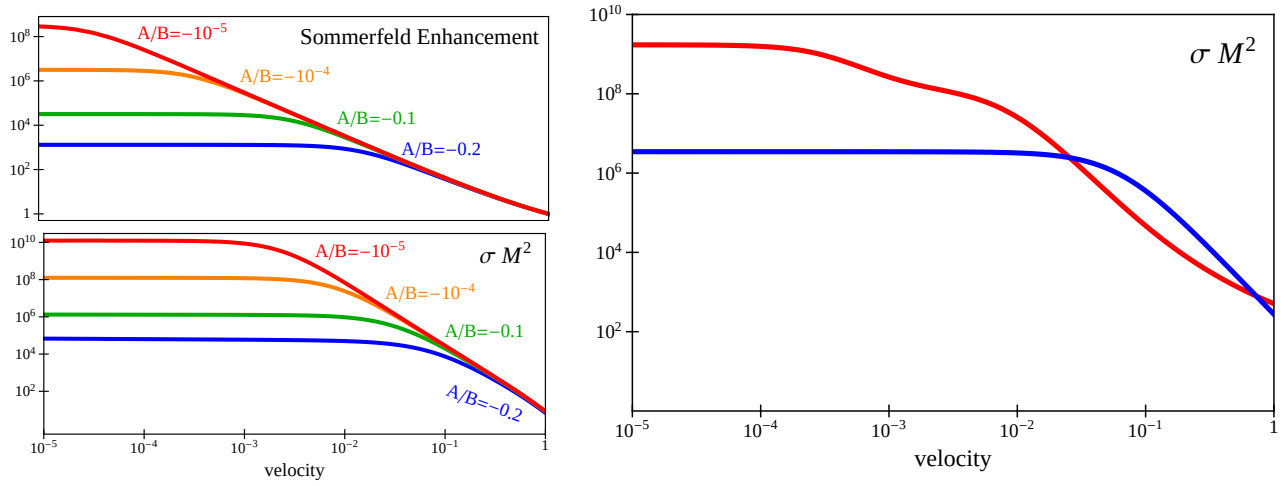

Figure 2.4: Core vs. cusp problem. LEFT: Sommerfeld enhancement (upper) and scattering cross section (lower) as a function of relative velocity for a range of low energy parameters A/B as discussed below (2.35) and $2\alpha M^2/f^2 = 1$. RIGHT: Total dark matter cross section as a function of velocity. Red: velocity dependent with $2\alpha M^2/f^2 = 34$, M = 5.8 TeV, $A/B = -10^{-3}$, and an additional short distance interaction, $M^2\sigma_{\text{short}}$ = 500. Blue: velocity independent cross section with no new short range interaction and $M =$ TeV, $2\alpha M^2/f^2 = 0.1$, $A/B = -6 \times 10^{-4}$.

enhanced cross section [85]. We leave a more thorough investigation of the astrophysical and cosmological bounds on the enhancement of singular potentials for future work.

As an example for how to apply Sommerfeld enhancement to address the dwarf galaxy scale astrophysical puzzles while simultaneously avoiding the bounds from galaxy cluster scale observations, we consider dark matter self interactions mediated by a light force carrier that generates a singular potential,

$$V(r) = \frac{-\alpha}{f^2}\frac{1}{r^3}. \tag{2.47}$$

The left side of Fig. 2.4 shows the Sommerfeld enhancement (upper) and the total cross section (lower) from such a model with a choice of parameters near the resonance. Observe that even for very small A/B, that is small $\cot\delta$ or large scattering phase, the cross section is saturated between the characteristic galaxy cluster velocities $v \sim 10^{-5}$ and dwarf galaxy velocities $v \sim 10^{-3}$. For $A/B \sim 10^{-5}$, as indicated by the red line in the lower figure,

this saturates to $\sigma/m_\chi \sim 10^{-2}$ cm^2/g for $m_\chi \sim$ TeV. This saturation occurs over the range of velocities where we would like a stronger velocity-dependence to avoid cluster scale bounds. In order to do this, we assume the existence of a short range interaction that contributes to the elastic scattering process with cross section $\sigma^{(0)}_{\text{short}}$. The long range mediators Sommerfeld enhance this cross section by the factor shown in the upper plot; observe that this enhancement decreases exponentially as one increases from dwarf galaxy velocities to galaxy cluster velocities. The total cross section is roughly (ignoring cross terms for simplicity),

$$\sigma_{\text{tot}}(v) \sim \sigma_{\text{elast}}(v) + S(v)\sigma^{(0)}_{\text{short}}. \tag{2.48}$$

Since the enhancement factors can be fairly large, the additional short range interaction can be weakly coupled, e.g. $\sigma^{(0)}_{\text{short}} M^2 \sim 10^4$ so that $\sigma_0 \sim 10^6$ pb for a TeV scale dark matter particle. The right side of Fig. 2.4 compares the velocity-dependence of this type of solution to another solution without enhanced short range physics. Fig. 2.5 shows contours of Sommerfeld enhancement as a function of velocity and elastic cross section, combining the data from the left-hand side of Fig. 2.4.

Finally, we remark on the use of Sommerfeld enhancement for generating indirect signals of dark matter through positrons and gamma rays [18, 41]. The excess of cosmic positrons observed by PAMELA [16] and later confirmed by FERMI [31] and AMS-02 [32] is a potential signal for dark matter annihilation. Since the cross section required to produce these signals is much larger than the required cross section for thermal relics, DM models that realize the positron excess typically require large Sommerfeld enhancements [86]. A study for non-singular dark sectors with Yukawa interactions was performed in [87, 51]; an investigation of how these bounds change for singular potentials is left for future work.

A recent speculative signal of indirect DM detection is the 135 GeV line in the FERMI gamma ray spectrum [15, 34, 35, 36, 37]. Indeed, gamma ray signatures were the original motivation for investigating Sommerfeld enhancement in dark matter [41]. The cross

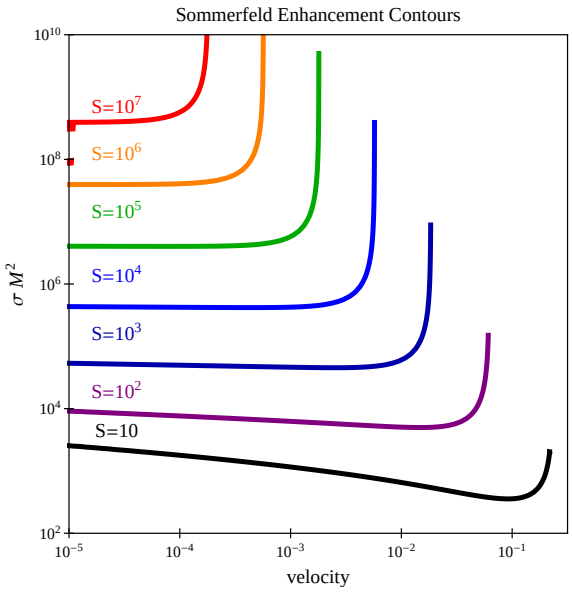

Figure 2.5: Contours of Sommerfeld enhancement from the singular potential (2.47) with $2\alpha M^2/f^2 = 1$ as a function of the DM velocity and elastic cross section σ.

section required for the line is about 10^{-27} cm^3/s which generically points toward a large boost factor, $S \approx 10^4$. It is possible to get such a large enhancement with a singular potential $V(r) = -\alpha/(fr^3)$, but since the dark matter mass must be 135 GeV this requires a low scale $f \approx 100$ GeV to avoid tuning in the UV. Dark matter models can generate such a feature, though these typically generate an unobserved continuum contribution to the spectrum [88]. Ways around difficulty were explored in [89, 90, 91, 92, 93].

2.7 Discussion

We have presented the effective non-relativistic theory of self-interacting dark matter parameterized to leading order in the relative velocity, v, and the exchanged momentum, q/Λ. The resulting potentials generically include singular terms which must be regulated and renormalized so that the resulting predictions are cutoff independent. We have shown

how this effective theory can be applied to calculate the Sommerfeld enhancement generated by singular potentials.

Using a simple toy model with a $1/r^3$ potential, we have found that on resonance one can generate enhancements as large as $S \sim 10^6$ at velocities on the order of $v \sim 10^{-3}$. This opens up promising directions for the astrophysical phenomenology of general self-interacting dark matter models. For example, extant astrophysical puzzles such as the core vs. cusp problem can be addressed with this velocity-dependent enhancement. A more thorough investigation and implications for specific UV models of these bounds is left for future work.

CHAPTER 3

WIMP DARK MATTER THROUGH THE DILATON PORTAL

Building a dark matter model often involves adding a number of free parameters which describe the coupling between dark matter and ordinary matter or between dark matter itself. These free parameters enlarge the parameter space of the model and therefore reduce the predictability of the model. In this chapter we avoid this problem by building a dark matter model where all the coupling constants are fixed by a symmetry called scale invariance. The spontaneous breaking of this symmetry gives rise to a new particle, called the dilaton, which mediate both dark matter self-interaction and interaction between dark matter and ordinary matter.

This chapter is based on the paper *WIMP Dark Matter through the Dilaton Portal* [94], written in collaboration with Kfir Blum, Csaba Csaki and Seung J. Lee.

3.1 Overview

We study a model in which dark matter couples to the Standard Model through a dilaton of a sector with spontaneously broken approximate scale invariance. Scale invariance fixes the dilaton couplings to the Standard Model and dark matter fields, leaving three main free parameters: the symmetry breaking scale f, the dilaton mass m_σ, and the dark matter mass m_χ. We analyze the experimental constraints on the parameter space from collider, direct and indirect detection experiments including the effect of Sommerfeld enhancement, and show that dilaton exchange provides a consistent, calculable framework for cold dark matter with f,m_σ,m_χ of roughly similar magnitude and in the range $\sim 1-10$ TeV. Direct and indirect detection experiments, notably future ground-based gamma ray and space-based cosmic ray measurements, can probe the model all the way to dark matter mass in the multi-TeV regime.

3.2 Introduction

Embedding the Standard Model (SM) partially or completely in a composite sector can solve the hierarchy problem, by making the Higgs boson composite. Often such a composite sector arises as the low-energy limit of an approximately scale invariant theory, where scale invariance is broken somewhere above the weak scale. If the breaking of scale invariance is spontaneous, then it is accompanied by a light dilaton (the Goldstone boson of spontaneously broken scale invariance) σ that couples to the fields in the composite sector through the trace of the energy-momentum tensor[1] of that sector [96, 97, 98, 99, 20, 100]

$$-\frac{\sigma}{f}\mathrm{Tr}T. \tag{3.1}$$

[1]To be precise the conserved, symmetric Callan-Coleman-Jackiw energy-momentum tensor [95] of the composite sector.

For massive particles, the coupling to σ is proportional to the particle masses, with the suppression scale f corresponding to the breaking of scale invariance. Similar couplings will be obtained to fields that mix the composite and elementary sector, except in that case the anomalous dimensions of the elementary fields will also appear in the formula. For a detailed discussion of the couplings of dilatons to various fields see Section 1.1.3 or [19].

The canonical dilaton Lagrangian (3.1) offers an economical way to couple the SM to new fields that could be singlets under the SM gauge symmetries and thus form an otherwise dark sector. In this paper we study the possibility that dark matter (DM) belongs to such dark sector and couples to the SM through Eq. (3.1). In the minimal set up that we explore here, three parameters determine the dynamics of thermal freeze-out in the early Universe: the breaking scale f, the dilaton mass m_σ, and the dark matter mass m_χ. Fixing one of these parameters such that the observed dark matter relic abundance is reproduced leaves a rather predictive framework. We show that a large parametric region exists where the solution is perturbative and produces cold, weakly interacting massive particle dark matter (WIMP), with $f,\, m_\sigma,\, m_\chi$ of roughly similar magnitude and in the range $\sim 1 - 10$ TeV.

Null results from dark matter direct detection experiments like LUX [101], XENON100 [102] and CDMS[103] put considerable pressure on WIMP models where DM couples to the SM through exchange of SM particles. The annihilation cross section $\sigma_{\rm ann} \sim 10^{-36}$ cm^2, required for WIMP relic abundance consistent with observations, is some ten orders of magnitude larger than the WIMP-nucleon elastic scattering cross section now probed by the direct detection experiments. This excludes Z boson exchange in all but fine-tuned corners of the parameter space, and requires some tuning for Higgs mediation as well. In contrast, the dilaton portal we analyze here quite generically evades the direct detection constraints in the bulk of the relevant parameter space, as the DM coupling to the SM resembles the case of Higgs exchange but with extra suppression of order $(v/f)^2\,(m_h/m_\sigma)^4$

with $v = \langle H \rangle = 246$ GeV and m_σ and the scale f automatically lying in the TeV ballpark to provide the correct relic abundance.

The idea that dark matter could couple to the SM via dilaton exchange was analyzed previously in Ref. [104] (where the dilaton was taken to be massless) and in Ref. [105] (for some specific warped extra dimensional models where the role of the dilaton was played by the radion). Our work generalizes the results of Ref. [105] and extends the analysis of [104] by adding the dilaton mass as a free parameter. This allows a more complete exploration of the parameter space and reveals effects such as Sommerfeld-enhanced annihilation. We also incorporate the most recent experimental bounds from direct and indirect detection as well as collider experiments.

The outline of this paper is as follows. In Sec. 3.3 we summarize the basic properties of the dilaton. We present its couplings, fixed mainly by the scale f with a few additional parameters characterizing the embedding of the SM matter into the composite sector, comment on expected NDA bounds on the dilaton mass, and present two benchmark models to be studied in the paper. Sec. 3.4 contains a calculation of the DM annihliation cross section due to dilaton exchange. After deriving a unitarity bound on the DM mass, we present the parameter space of the theory where the observed relic abundance is reproduced. In Sec. 3.5 we compare the DM-nucleon scattering cross sections to the experimental bounds from the latest round of direct detection measurements, finding that large regions of the parameter space are compatible with the bounds. In Sec. 3.6 we consider constraints from indirect detection of gamma rays and cosmic ray antiprotons. We conclude in Sec. 3.7. App. A summarizes collider bounds on the dilaton, considering LEP, Tevatron and the LHC. App. B contains cross-section formulae for the sub-leading annihilation channels that we omit in the body of the text for clarity.

3.3 The Dilaton Mediated Dark Matter Model

We start by considering the effective theory describing an approximately scale invariant sector with scale invariance spontaneously broken at the scale f. The Goldstone boson corresponding to this breaking, called the dilaton $\widetilde{\sigma}(x)$, can be parametrized via a spurion field $\Phi(x)$ as [19]

$$\Phi(x) \equiv f e^{\widetilde{\sigma}(x)/f} \tag{3.2}$$

such that under a scale transformation $x \to xe^{\lambda}$ we have $\Phi(x) \to e^{\lambda}\Phi(e^{\lambda}x)$ and $\langle\Phi\rangle = f$. This Φ is not (necessarily) an elementary scalar - rather it is the analog of the Σ field of non-linear σ-models for the case of spontaneously broken scale invariance, which allows one to construct the low-energy effective Lagrangian for the dilaton that captures the essential features of the broken symmetry. To obtain a canonically normalized dilaton kinetic term it is convenient to do a field redefinition such that $\Phi(x) = \sigma(x) + f$ [20] where σ is now the canonically normalized dilaton field. Using a spurion analysis one can then find the low energy theory below the cutoff scale $4\pi f$ by inserting powers of Φ/f in the SM Lagrangian to make it scale invariant. After electroweak symmetry breaking one finds the following effective action describing the interactions of the canonically normalized dilaton with the SM fields [99, 20, 19]

$$\begin{aligned}
\mathcal{L}_\sigma &= \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - \frac{1}{2}m_\sigma^2\sigma^2 - \frac{5}{6}\frac{m_\sigma^2}{f}\sigma^3 - \frac{11}{24}\frac{m_\sigma^2}{f^2}\sigma^4 + \ldots - \left(\frac{\sigma}{f}\right)\left[\sum_\psi (1+\gamma_\psi) m_\psi \bar{\psi}\psi\right] + \\
&+ \left(\frac{2\sigma}{f} + \frac{\sigma^2}{f^2}\right)\left[m_W^2 W^{+\mu}W_\mu^- + \frac{1}{2}m_Z^2 Z^\mu Z_\mu - \frac{1}{2}m_h^2 h^2\right] + \frac{\alpha_{\rm EM}}{8\pi f}c_{\rm EM}\sigma F_{\mu\nu}F^{\mu\nu} + \\
&+ \frac{\alpha_{\rm s}}{8\pi f}c_{\rm G}\sigma G_{a\mu\nu}G^{a\mu\nu}.
\end{aligned} \tag{3.3}$$

The sum on ψ runs over the SM fermions, which are assumed to be partially composite with light fermions being mainly elementary and the top quark mainly composite. γ_ψ corresponds to the anomalous dimension of fermionic operators responsible for generating the SM fermion masses after mixing between the elementary and composite sectors. For

composite fermions the anomalous dimension is expected to be small $\gamma_\psi \simeq 0$, while for light fermions the anomalous dimension may be sizable.

Naive dimensional analysis (NDA) limits the plausible size of the dilaton mass. For example, considering the dilaton self-energy loop from the trilinear coupling in Eq. (3.3) we find that

$$m_\sigma \leq 4\pi f \tag{3.4}$$

to ensure that the one-loop correction of the dilaton mass remains below the tree-level mass, and that the couplings of the dilaton to matter remain under control [106]. This is just the reflection of the fact that this theory has an intrinsic cutoff of order $\Lambda \sim 4\pi f$, and we should treat it as an effective theory valid below that scale. Note also that it is difficult to make the dilaton much lighter than the cutoff scale. In generic models there is a tuning of order $\frac{m_\sigma}{\Lambda}$ necessary to lower the dilaton mass [19, 106, 107], though special constructions can potentially alleviate this tuning [108, 109]. We will require that the dilaton is not lighter than $f/10$.[2]

A few additional comments are in order about Eq. (3.3). The cubic and quartic dilaton self interaction arise from expanding the effective dilaton potential which includes a scale invariant term, Φ^4, and small explicit sources of scale symmetry breaking such as $\Phi^{4-\epsilon}$. These two terms are expected to be the dominant ones in the effective potential of the non-linear field Φ, and is applicable to all types of theories with spontaneous breaking of scale invariance, irrespective of whether there is an elementary scalar breaking the symmetry, or the breaking happens dynamically. The $\Phi^{4-\epsilon}$ term represents the effect of small explicit breaking of scale invariance. The effect of any explicit breaking term will be to generate a term in the effective potential with power different from 4. If the breaking is small,

[2]Very light dilatons (while generically tuned) can have interesting phenomenology. For example it could be a dark matter component itself, or lead to a form of dark radiation. However in that scenario several additional important cosmological and astrophysical constraints have to be satisfied, including overclosure of the Universe, star and supernova cooling and fifth force constraints. An exploration of these constraints and the viability of the very light dilaton mass region is beyond the scope of this work.

the power is expected to lie close to 4, hence the usual choice of power $4-\epsilon$, where ϵ usually represents the anomalous dimension of an explicitly non-scale invariant operator. We want to emphasize again that this parametrization of the potential is expected to be quite general, covering both cases of breaking via an elementary scalar and dynamical breakings of scale invariance. Requiring that $\langle\Phi\rangle = f$ and that $\frac{d^2V(\Phi)}{d\Phi^2} = m_\sigma^2$ fixes the parameters of the dilaton potential. The full expression for the cubic self-coupling of the dilaton is $\frac{5-\epsilon}{6}\frac{m_\sigma^2}{f}$, which for small ϵ we can approximate as $\frac{5}{6}\frac{m_\sigma^2}{f}$[3], while for small ϵ the quartic is $11/24\frac{m_\sigma^2}{f^2}$. Away from the $\epsilon \to 0$ limit, the prefactor of the cubic coupling can lie anywhere in the interval $[2/6, 5/6]$ [20]. For simplicity, throughout this paper we have used the limiting value $5/6$ for the cubic, though we have verified that this does not influence our results significantly. The coupling of the dilaton to massless gauge bosons arises from two sources; just like for the SM Higgs, the dilaton receives a contribution from top quark and W boson loops, but in addition there is a direct contribution from the trace anomaly. The trace anomaly is proportional to the β-functions: the actual contribution will be the difference between the β-function above and below the symmetry breaking scale. Thus this contribution depends on the details of what fraction of the composite sector is actually charged under the unbroken SM gauge symmetries, and what fraction of the SM fields are composites. For example, the coupling to gluons c_G receives a contribution from the trace anomaly and from a top loop and is given by

$$c_G = b^{(3)}_{\text{IR}} - b^{(3)}_{\text{UV}} + \frac{1}{2}F_{1/2}(x_t) \tag{3.5}$$

where $b^{(3)}_{UV,IR}$ are the QCD β-function coefficients above and below the scale f. This is a free parameter of the theory, which gives a measure of the QCD charges of the scale invariant sector. The function $F_{1/2}$ is the usual triangle diagram contribution of a fermion

[3]Note that for $\epsilon = 0$ both the mass m_σ^2 and the cubic coupling vanish. Here we are assuming that ϵ is small but non-vanishing.

given by

$$F_{1/2}(x) = 2x\left[1+(1-x)f(x)\right] \tag{3.6}$$

$$f(x) = \begin{cases} \left[\sin^{-1}\left(1/\sqrt{x}\right)\right]^2 & x \geq 1 \\ -\frac{1}{4}\left[\log\left(\frac{1+\sqrt{x-1}}{1-\sqrt{x-1}}\right) - i\pi\right]^2 & x < 1 \end{cases} \tag{3.7}$$

where $x_t = 4m_t^2/m_\sigma^2$ [99, 19]. A similar expression applies to the coupling to photons.

Some of the results in the following sections (in particular the direct detection and collider signals) depend on the parameters c_G and c_{EM}. To this end we define two benchmark model examples which we will study in detail.

Model A: This is the well-studied case proposed in [20] where the entire SM is composite, corresponding to $b_{UV} = 0, b_{IR} = b_{SM}$, giving rise to the parameters $b^3_{UV} - b^3_{IR} = -7,\ b^{EM}_{UV} - b^{EM}_{IR} = 11/3$. Note that for a light dilaton these b's depend somewhat on the dilaton mass: for example $b^3_{UV} - b^3_{IR} = -11 + 2n/3$, with n denoting the number of quarks whose mass is smaller than $m_\sigma/2$.

Model B: This is a limit of the well-motivated case when only the right-handed top and the Goldstone bosons needed for electroweak symmetry breaking are composites, while we minimize the β-functions of the UV to be as small as possible, resulting in $b^3_{UV} = b^{EM}_{UV} = 0,\ b^3_{IR} = -1/3,\ b^{EM}_{IR} = -11/9$. Note however that b_{UV} is in fact a free parameter depending on the actual UV theory, and its value here has been chosen only for illustration.

The final ingredient of the model is χ, the dark matter particle, which can be spin 0, 1/2 or 1. We assume that χ is a composite of the conformal sector, and does not have any direct coupling to the standard model fields which are mainly elementary. The couplings of χ to the dilaton are fixed by a spurion analysis and follow the rules of couplings of

generic massive composites[104]:

$$\mathcal{L}_{\rm DM} \supset \begin{cases} -\left(1+\frac{2\sigma}{f}+\frac{\sigma^2}{f^2}\right)\frac{1}{2}m_\chi^2\chi^2 & \text{Scalar} \\ -\left(1+\frac{\sigma}{f}\right)m_\chi\bar{\chi}\chi & \text{Fermion} \\ \left(1+\frac{2\sigma}{f}+\frac{\sigma^2}{f^2}\right)\frac{1}{2}m_\chi^2\chi_\mu\chi^\mu & \text{Gauge boson.} \end{cases} \tag{3.8}$$

For simplicity we assume that a $\mathcal{Z}_2$ symmetry renders χ to be a stable particle. For the fermionic case, we assume that χ is a Dirac fermion.

3.4 Relic Abundance

In this section we present the calculation of the relic abundance of the dark matter field χ, where annihilations into SM states are assumed to proceed via dilaton exchange, and exhibit the relevant parameter space of the theory. As usual, for small relative velocities v the velocity-weighted annihilation cross section can be expanded as $\sigma v = a + bv^2$. At the freeze-out temperature T_F we have $\langle v^2\rangle = 6/x_F$ where $x_F = m_\chi/T_F$. The value of x_F can then be determined by solving the Boltzmann equation in an expanding Universe:

$$x_F = \ln\left(\frac{5}{4}\sqrt{\frac{45}{8}}\frac{g}{2\pi^3}\frac{M_{\rm Pl}m_\chi(a+6b/x_F)}{\sqrt{g^*}\sqrt{x_F}}\right), \tag{3.9}$$

where g is the number of degrees of freedom of the dark matter particle and g^* is the effective number of relativistic degrees of freedom in thermal equilibrium during dark matter freeze-out. Once x_F is determined the dark matter relic abundance is given by

$$\Omega_\chi h^2 \approx \frac{1.07\times 10^9}{\text{GeV}M_{\rm Pl}\sqrt{g^*}}\frac{x_F}{a+3(b-a/4)/x_F}. \tag{3.10}$$

As we show below, the dark matter annihilation cross section (and thus the parameters a, b) in the model considered here is calculated in terms of m_χ, m_σ and f. Requiring that the observed relic abundance $\Omega_\chi h^2 = 0.1199 \pm 0.0027$ [10] is reproduced will thus impose one non-trivial relation and reduce the parameter space of the model. Next we map out

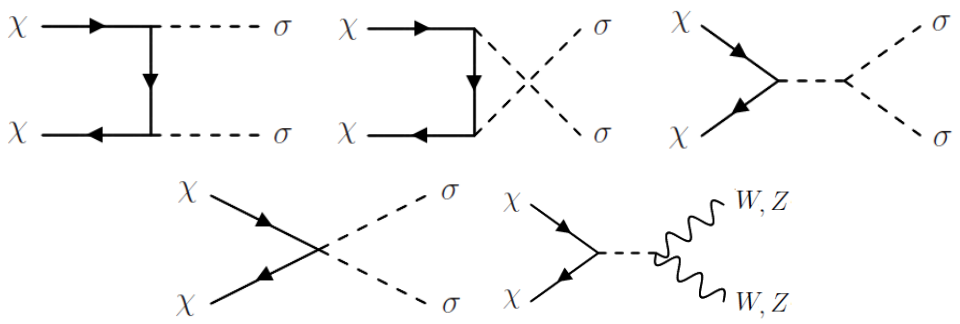

Figure 3.1: Leading annihilation diagrams of dark matter in the regime $m_\chi \gg m_t$. For fermionic dark matter there is no direct annihilation to dilatons.

this relation in detail, obtaining the reduced parameter space of the theory to be tested by direct and indirect detection experiments as well as collider searches.

3.4.1 Annihilation cross sections

The dominant dark matter annihilation channels for $m_\chi \gg m_t$ are $\chi\chi \to \sigma\sigma, WW, ZZ$, shown in Fig. 3.1. The dominant channels contain factors of m_χ/f, to be compared with all other sub-leading channels (for example s-channel dilaton exchange with quark or higgs final states) that contain factors of m_q/f or m_h/f instead and are thus suppressed by relative powers of $m_{q,h}/m_\chi$. Below we present analytical expressions for the dominant channels in the limit $m_\chi \gg m_t$, for the cases of scalar, fermion and vector dark matter. Formulae for the sub-leading annihilation channels can be found in Appendix E. For numerical results all of the allowed annihilation channels are included.

Scalar dark matter

Scalar dark matter annihilation is dominated by s-wave processes. The approximate expressions of the cross sections are

$$\sigma v(\chi\chi \to WW) \simeq \frac{m_\chi m_W^4 \sqrt{m_\chi^2 - m_W^2}\left(2 + \frac{(2m_\chi^2 - m_W^2)^2}{m_W^4}\right)}{2\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2}, \tag{3.11}$$

$$\sigma v(\chi\chi \to \sigma\sigma) \simeq \frac{m_\chi \sqrt{m_\chi^2 - m_\sigma^2}|2(2m_\chi^2 - m_\sigma^2)^2 + 2m_\sigma^4 - i\text{Im}\left(\Pi(4m_\chi^2)\right)\left(2m_\chi^2 + m_\sigma^2\right)|^2}{16\pi f^4 (2m_\chi^2 - m_\sigma^2)^2 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2}. \tag{3.12}$$

Note that the second term in the parenthesis of Eq. (3.11), corresponding to the formation of longitudinal gauge boson modes, becomes proportional to m_χ^4/m_W^4 in the limit $m_\chi \gg m_W$. In this limit, the m_W^4 pre-factor is cancelled such that the overall cross section scales like m_χ^2/f^4.

In the expressions above $\Pi(p^2)$ is the 1PI self-energy insertion for the dilaton, which on-shell is related to the width via $m_\sigma \Gamma_\sigma = -\text{Im}\left(\Pi(m_\sigma^2)\right)$. Note that we only include the imaginary part in our calculations. The real part (once properly renormalized) is expected to be a moderate correction to the existing real part of the propagator, which will result in small shifts to the precise shape of the contours presented below, but can not qualitatively change the results, as long as the NDA bound (3.4) on the dilaton mass is obeyed. On the other hand properly incorporating the non-vanishing imaginary part can give significant shifts in the resulting cross sections especially close to the resonance.

The total width of the dilaton is the sum of the partial widths to Higgs, quarks, massive gauge bosons and dark matter, which in the limit $m_\sigma \gg m_t$ is dominated by the decays to massive gauge bosons

$$\Gamma_\sigma(\sigma \to WW) = \frac{m_W^4}{4\pi m_\sigma f^2}\sqrt{1 - 4\frac{m_W^2}{m_\sigma^2}}\left(2 + \frac{(m_\sigma^2 - 2m_W^2)^2}{4m_W^4}\right). \tag{3.13}$$

The processes $\chi\chi \to ZZ$ and $\sigma \to ZZ$ are obtained from Eqs. (3.11-3.13) by replacing m_W

by m_Z and dividing by 2 to account for the phase space of identical final state particles. In Appendix E.1 we collect the contributions of the other channels to the dilaton decay width.

Fermionic dark matter

For fermionic dark matter, the annihilation channels have no s-wave contribution, thus the dominant contribution is a p-wave process which is suppressed by a factor of v^2. We find

$$\sigma v(\chi\bar{\chi} \to WW) \simeq v^2 \frac{m_\chi m_W^4 \sqrt{m_\chi^2 - m_W^2}\left(2 + \frac{(2m_\chi^2 - m_W^2)^2}{m_W^4}\right)}{16\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{3.14}$$

$$\begin{aligned}\sigma v(\chi\bar{\chi} \to \sigma\sigma) \simeq\ & v^2 \Bigg[\frac{m_\chi^5 \sqrt{m_\chi^2 - m_\sigma^2}\left(9m_\chi^4 - 8m_\sigma^2 m_\chi^2 + 2m_\sigma^4\right)}{24\pi f^4\left(16m_\chi^8 - 32m_\chi^6 m_\sigma^2 + 24m_\chi^4 m_\sigma^4 - 8m_\sigma^6 m_\chi^2 + m_\sigma^8\right)} \\ &+ \frac{25 m_\chi m_\sigma^4 \sqrt{m_\chi^2 - m_\sigma^2}}{128\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \\ &- \frac{5 m_\chi^3 m_\sigma^2 \sqrt{m_\chi^2 - m_\sigma^2}(5m_\chi^2 - 2m_\sigma^2)}{48\pi f^4\left(4m_\chi^4 - 4m_\sigma^2 m_\chi^2 + m_\sigma^4\right)} \text{Re}\left(\frac{1}{4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)}\right)\Bigg].\end{aligned} \tag{3.15}$$

Vector dark matter

For vector boson dark matter the annihilation is again dominated by s-wave processes:

$$\sigma v(\chi\chi \to WW) \simeq \frac{m_\chi m_W^4 \sqrt{m_\chi^2 - m_W^2}\left(2 + \frac{(2m_\chi^2 - m_W^2)^2}{m_W^4}\right)}{6\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{3.16}$$

$$\begin{aligned}\sigma v(\chi\chi \to \sigma\sigma) \simeq\ & \frac{m_\chi \sqrt{m_\chi^2 - m_\sigma^2}}{144\pi f^4 (2m_\chi^2 - m_\sigma^2)^2 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2}\Big(708 m_\chi^8 \\ &+ 44 m_\sigma^2 m_\chi^2 \text{Im}^2\left(\Pi(4m_\chi^2)\right) - 28 m_\sigma^4 \text{Im}^2\left(\Pi(4m_\chi^2)\right) - 1600 m_\sigma^2 m_\chi^6 \\ &+ 1424 m_\chi^4 m_\sigma^4 - 576 m_\sigma^6 m_\chi^2 + 11 m_\sigma^4 \text{Im}^2\left(\Pi(4m_\chi^2)\right) + 96 m_\sigma^8\Big).\end{aligned} \tag{3.17}$$

3.4.2 Unitarity considerations

We emphasize again that the WW and ZZ annihilation channels are important because of the enhanced contributions of the longitudinal modes. Note that Ref. [104] neglected these channels due to the suppression of the W/Z couplings by $m_{W,Z}/f$. However as we have shown in the previous section, these factors are cancelled in the limit $m_\chi \gg m_Z$ due to the contributions of the longitudinal modes which grow with the CM energy/dark matter mass.

For large DM mass, the gauge boson longitudinal modes might violate unitarity. This is analogous to the unitarity violation in elastic WW scattering in the standard model without the Higgs. However here the Higgs does not save unitarity. Thus we will have a unitarity bound on the DM mass, related to the built-in cutoff for the theory above which it is expected to be strongly coupled. One can estimate the unitarity bound on m_χ by considering the contribution of the longitudinal mode to the scattering amplitude in the large DM mass limit, given by $\mathcal{M} \approx 2m_\chi^2/f^2$ for either scalar, fermion or vector DM. The resulting s-wave partial wave amplitude $a_0 \approx m_\chi^2/(16\pi f^2)$ satisfies the unitarity bound $|\Re(a_0)| \leq 1/2$ if

$$m_\chi \leq \sqrt{8\pi} f. \tag{3.18}$$

This unitarity bound on m_χ is slightly more constraining than the NDA estimate for the cutoff $m_\chi \lesssim \Lambda_{NDA} = 4\pi f$. A similar analysis for the annihilation to dilatons results in the same upper bound.

3.4.3 The basic parameter space

We now analyze the parameter space of the model that is compatible with the observed dark matter relic density. Fig. 3.2 shows the available parameter space where the observed

relic density can be reproduced by an appropriate choice of the symmetry breaking scale f. The top left, top right, and bottom panels show the results for scalar, fermion, and vector dark matter, respectively. The x- and y-axes correspond to the dilaton and dark matter mass, while the contours show the value of f that is required to obtain the observed dark matter relic density.

For concreteness, in the rest of this section we discuss the scalar dark matter case. We later summarize the results for fermion and vector dark matter. To understand the results shown in Fig. 3.2 (top left, as we are focusing on the scalar example), we consider the different parametric regions in turn. Consider the case $m_\chi, m_\sigma \gg m_Z$, where annihilation to WW, ZZ and, if kinematically allowed, $\sigma\sigma$ dominates. Assume first $m_\chi > m_\sigma$, corresponding to the upper-left region in Fig. 3.2. Here we have

$$\langle \sigma v \rangle \approx \frac{m_\chi^2}{4\pi f^4} \approx 3 \times 10^{-26} \left(\frac{f}{6\ \mathrm{TeV}}\right)^{-2} \left(\frac{m_\chi}{f}\right)^2 \mathrm{cm}^3/\mathrm{s} \qquad (\text{valid for } \mathrm{m}_\chi \gg \mathrm{m}_\sigma). \quad (3.19)$$

Recall that relic abundance consistent with observations requires $\langle \sigma v \rangle \approx 3 \times 10^{-26}$ cm^3/s, and that $\Omega_\chi h^2 \propto \langle \sigma v \rangle^{-1}$, imposing the relation $m_\chi = f^2/(6\mathrm{TeV})$. Combining this with the unitarity bound $m_\chi \sim \sqrt{8\pi} f$ obtained above, we find an upper bound $f < 30$ TeV. Violating this bound leads to DM annihilation cross section that is too small, and so DM relic density that is too high to match observations. A caveat in this derivation is that our dark matter particle may co-annihilate with extra particles in the dark sector. If this co-annihilation is efficient, due to some mass degeneracy in the dark sector and large cross sections, then it would relax the bound on f, allowing f to be somewhat larger than 30 TeV. Even taking this caveat into account, a rough bound $f \lesssim 100$ TeV is still expected to hold. We note that this derivation of the bound on f is compatible with the unitarity argument of [2], that showed that $m_\chi \lesssim 100$ TeV is required in general from S-matrix unitarity (we update their early result here by using the currently measured DM relic density). Plugging the model-independent upper bound on m_χ from Ref. [2] into Eq. (3.19), we obtain again $f \lesssim 30$ TeV. The consistency between Eq. (3.18) and the unitarity bound of [2] implies that

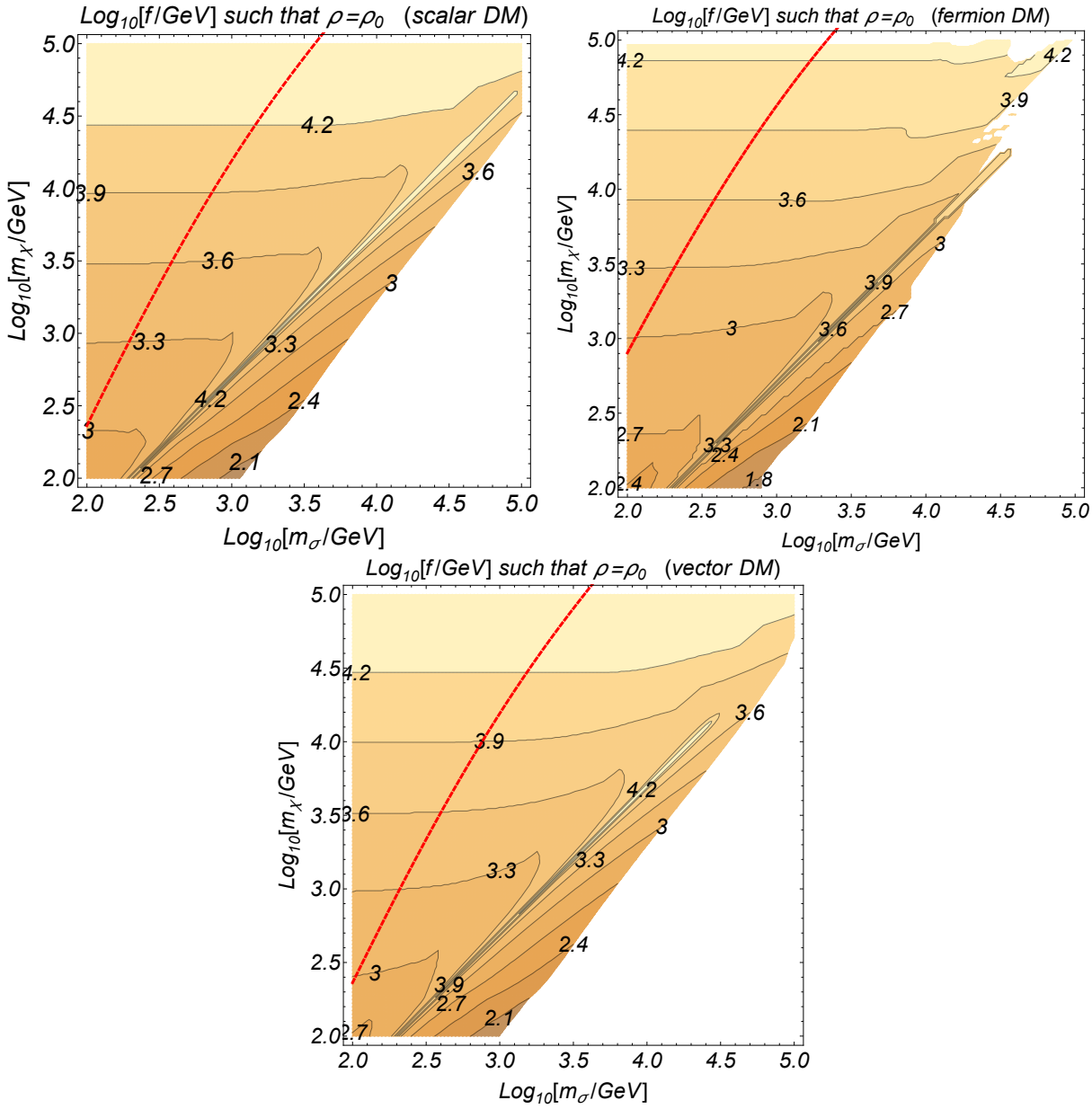

Figure 3.2: Parameter space for scalar (top left), fermion (top right) and vector (bottom) dark matter with freeze-out mediated by dilaton exchange. The x and y axes correspond to the dilaton and dark matter mass, respectively. Contours show the value of the symmetry breaking scale f, that is required in order to obtain the observed dark matter relic density. In the blank region in the lower-right part of the plot, there is no real solution for f that provides the correct relic density while satisfying Eq. (3.4). Above the red dashed line $m_\sigma < f/10$, signaling some degree of fine-tuning. Note that the model-independent unitarity bound of Ref. [2] implies $m_\chi \lesssim 10^5$ GeV (see text).

Eq. (3.18) is satisfied throughout the parameter space shown in Fig. 3.2.

Next, consider the region with $m_\sigma \gg m_\chi$, so that the $\chi\chi \to \sigma\sigma$ channel is kinematically forbidden. This region corresponds to the lower-right part of Fig. 3.2. In this regime, and still assuming $m_\chi \gg m_Z$, one finds the following approximate form for the cross section:

$$\langle \sigma v \rangle \sim \frac{3 m_\chi^6}{\pi f^4 m_\sigma^4} \approx 2 \cdot 10^{-26} \left(\frac{m_\chi}{350\ \mathrm{GeV}}\right)^6 \left(\frac{\mathrm{TeV}}{f}\right)^4 \left(\frac{\mathrm{TeV}}{m_\sigma}\right)^4 \mathrm{cm}^3/\mathrm{s} \quad (m_\chi \gg m_W,\ m_\sigma \gg m_\chi). \tag{3.20}$$

As one increases the dilaton mass m_σ the symmetry breaking scale f needs to decrease in order to keep the relic abundance fixed. However, one will very quickly need to lower f below the value $m_\sigma/4\pi$, implying that we have left the regime of validity of our effective theory. Therefore most of the lower left region will be excluded based on this criterion. Of course the exact shape of the excluded region will be somewhat uncertain: it depends on the exact onset of strong coupling, and can also be slightly modified by strong co-annihilations in the dark sector. Nevertheless, even in this case we expect that the allowed region would shift only slightly.

The resonance at $m_\sigma = 2m_\chi$ is clearly visible in Fig. 3.2. The approximate expression of the cross section close to the resonance region is

$$\langle \sigma v \rangle \sim \frac{3 m_\chi^6}{\pi \left[(\Delta m)^4 f^4 + \frac{9 m_\chi^8}{4\pi^2} \right]}\ , \tag{3.21}$$

where $\Delta m^2 = 4m_\chi^2 - m_\sigma^2$, measuring the deviation from the exact location of the resonance. In this region (but above the blank region corresponding to Eq. (3.20)), a large value of f is required to reduce the otherwise too high annihilation cross section. Note, that once $m_\chi \sim 40$ TeV the cross section falls below the observed value even without a contribution from the resonance. Above those masses one does not expect any more resonant behavior, which is indeed what is reflected in Fig. 3.2. We note that numerical resolution affects the size of f that is displayed in Fig. 3.2 exactly on the resonance line, as $f \to \infty$ for $\Delta m^2 \to 0$. Of course, living exactly on resonance corresponds to an extremely fine-tuned

parametric set-up. Note that beyond the mere parametric fine-tuning, another issue here is that $f \gg m_\sigma$ would imply dynamical fine-tuning as well.

We conclude the discussion of the scalar DM case by considering the scenario proposed in Ref. [19], that entertained the possibility of having the newly discovered Higgs-like particle itself be the dilaton. For the dilaton to mimic the Higgs, we must have $m_\sigma \approx m_h = 126$ GeV and $f \approx v = 246.2$ GeV. For these values of m_σ and f, we find that the dark matter mass that is needed for correct relic abundance is $m_\chi \approx 52$ GeV if the dark matter is a scalar. The leading annihilation channels at this value of m_χ are to bottom and charm quarks and tau leptons. Larger values of m_χ result in relic abundance that is too low, while lower values of m_χ give a too-high relic abundance. This means that $m_\chi \approx 52$ GeV is an upper bound for scalar dark matter mass in our framework in the Higgs-like dilaton scenario. As we show in Sec. 3.5, such a low scalar dark matter mass is excluded by direct detection limits. Similar results are obtained for the case of fermion and vector DM, as presented in the second and third plots in Fig. 3.2. The higgs-like dilaton scenario would require fermion dark matter of 61 GeV, or vector DM of 56 GeV. As we will see both of these cases are excluded by the direct detection bounds.

Finally, note that in part of the parameter space depicted in Fig. 3.2 the DM annihilation cross section receives large non-perturbative corrections at low center of mass velocities (Sommerfeld enhancement). In our model, at large DM mass when the effective coupling m_χ/f is not far from the perturbativity limit, the effect induces a sizable correction to the relic abundance calculation. We compute the Sommerfeld enhancement in Sec. 3.6 and include it in a simplified form in the calculation of Fig. 3.2, by rescaling the tree-level annihilation cross section by the Sommerfeld enhancement factor at relative DM velocity $v = 0.3$, corresponding roughly to the thermal freeze-out kinematics. In most of the parameter space, corresponding to perturbative coupling $(m_\chi/f)^2/4\pi \ll 1$, the correction to

the derived value of $f(m_\chi, m_\sigma)$ fixed by the relic abundance requirement is insignificant [4].

3.5 Direct Detection

Having defined the parameter space of the theory that reproduces the correct relic abundance, we now study direct detection constraints. For direct detection we need to consider the elastic cross section of a dark matter particle that scatters off a nucleon. The dilaton interacts with quarks q and the gluons $G^{a\mu\nu}$ inside a nucleon [110, 111]. Thus the relevant part of the dilaton effective Lagrangian is

$$\mathcal{L} \supset -\sum_q \frac{\sigma}{f}(1+\gamma_q) m_q q\bar{q} + \frac{\alpha_s}{8\pi f} c_G G^2 \,. \tag{3.22}$$

To estimate the anomalous dimension for quarks, one can consider the corresponding warped extra dimensional models where the anomalous dimension is determined [99] by $1+\gamma = c_L - c_R$, where $c_{L,R}$ are the bulk fermion mass parameters. For typical warped fermion scenarios we find for example $\gamma_s \sim 0.16$, which we neglect in the bounds below.

Taking the matrix element between nucleon states yields the effective nucleon-dilaton Lagrangian

$$\mathcal{L}_{\sigma nn} \;=\; y_n \sigma n \bar{n} \tag{3.23}$$

where the coefficient y_n is determined by the f_q^n, R^n hadronic matrix elements:

$$y_n \equiv -\sum_q f_q^n \frac{m_n}{f} + R^n \frac{c_G}{8\pi f} \,. \tag{3.24}$$

[4] In fine-tuned regions of the parameter space, where the Sommerfeld effect hits a resonance, DM annihilation re-coupling can significantly affect the relic abundance calculation [87]. We ignore this effect here and comment about it in Sec. 3.6.

For these matrix elements we use the values from [111, 112, 113]:

$$\begin{aligned} f_q^n &= \langle n|\bar{q}q|n\rangle \frac{m_q}{m_n} \\ f_u^n &\simeq f_d^n \simeq 0.022 \\ f_s^n &\simeq 0.043 \\ f_c^n &\simeq 0.0814 \\ f_b^n &\simeq 0.0785 \\ f_t^n &\simeq 0.0820 \\ R^n &= \alpha_s \langle n|G^a_{\mu\nu} G^{a\mu\nu}|n\rangle \simeq -2.4\text{GeV} \end{aligned} \tag{3.25}$$

With this effective interaction the scattering cross section between dark matter and nucleons is given by

$$\sigma_{\chi,n} \approx \frac{y_n^2}{\pi} \left(\frac{m_\chi}{f}\right)^2 \frac{m_n^2}{m_\sigma^4} \tag{3.26}$$

for either scalar, fermionic or vector dark matter.

Fixing the scale f for given m_σ and m_χ to match the relic abundance, we plot the DM-nucleon elastic scattering cross section as a function of the dark matter mass for a few dilaton mass values. The results are illustrated on Fig. 3.3 along with the recent direct detection constraints from the LUX experiment [101]. We have also included the effects of the collider bounds on the dilaton from the LHC and other machines (see Appendix D). These plots show that most of the parameter space is currently allowed both by the dark matter direct detection experiments and also by the collider constraints, as long as $m_\sigma \gtrsim 200$ GeV.

As discussed in Sec. 3.4.3, for $m_\chi \gg m_t$ and away from the resonance the annihilation cross section is proportional to m_χ^2/f^4. Moreover, since $y_n \propto 1/f$, we can see that the elastic scattering cross section is proportional to the same combination m_χ^2/f^4. Thus in the appropriate regime the elastic cross section will be independent of the dark matter mass, as can be seen in Fig. 3.3.

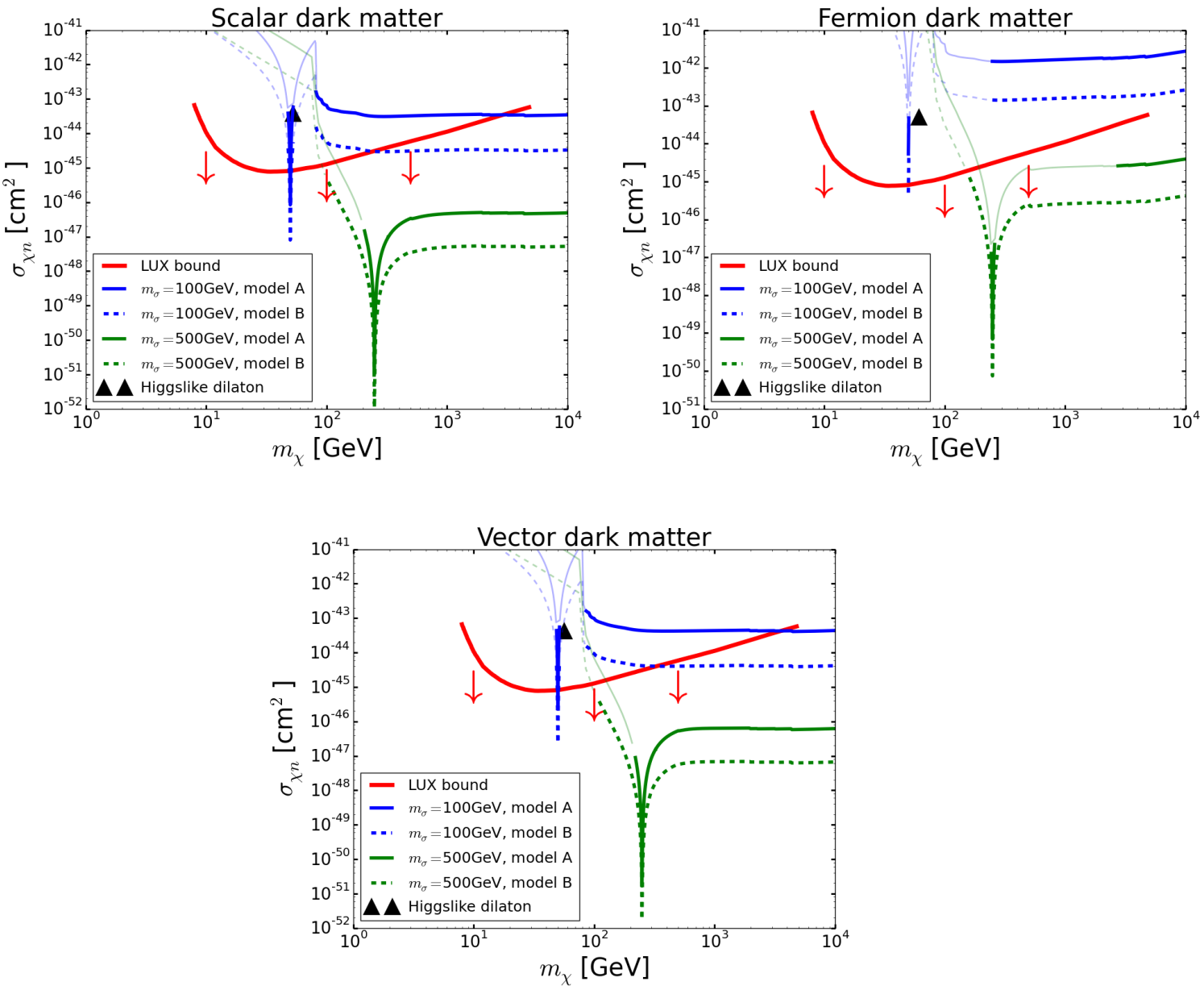

Figure 3.3: Nucleon-dark matter elastic cross section as a function of dark matter mass. The red arrows point towards the non-excluded region. The lighter portion of the curves are already excluded by bounds from collider experiments searching for a dilaton.

3.6 Sommerfeld Enhancement and Indirect Detection

We now consider the prospects for indirect detection of dark matter annihilation via gamma ray and cosmic ray antiproton flux measurements [5]. We limit the discussion to the case in which the DM χ is a real scalar field. We expect similar results for the vector DM case; the fermion DM case will not have significant cosmic ray signatures as its annihilation is p-wave suppressed in the small virial velocity of the Milky Way and its dwarf satellite galaxies.

The parameter space of interest for the model includes the regime where $m_\chi > m_\sigma$. In this regime, dilaton exchange produces an attractive Yukawa potential $-\frac{\alpha}{r}e^{-m_\sigma r}$, with $\alpha = \frac{m_\chi^2}{4\pi f^2}$, that affects the dark matter annihilation process giving rise to Sommerfeld enhancement (SE; see e.g. [115, 41]) that needs to be taken into account in the indirect detection estimates. In the top panel of Fig. 3.4 we plot the effective SE factor (denoted SE_{eff}) in the $\{m_\sigma, m_\chi\}$ plane, fixing the value of the scale f at each point to match the observed dark matter relic abundance. We define SE_{eff} as the value of the SE today in the Galactic halo, normalized to its value during DM freeze-out when $v \sim 0.3$. In our calculation we use an approximate formula for the SE factor [40, 87, 116],

$$SE \approx \frac{\pi}{\epsilon_v} \frac{\sinh\left(\frac{12\epsilon_v}{\pi\epsilon_\phi}\right)}{\cosh\left(\frac{12\epsilon_v}{\pi\epsilon_\phi}\right) - \cos\left[2\pi\sqrt{\frac{6}{\pi^2\epsilon_\phi} - \left(\frac{12\epsilon_v}{\pi\epsilon_\phi}\right)^2}\right]}, \tag{3.27}$$

where $\epsilon_v \equiv \frac{v}{2\alpha} = \frac{2\pi v f^2}{m_\chi^2}$ and $\epsilon_\phi \equiv \frac{m_\sigma}{\alpha m_\chi} = \frac{4\pi m_\sigma f^2}{m_\chi^3}$. We set the value of the dark matter particles' relative velocity to $v = 10^{-3}$, appropriate for annihilation in the Galactic halo. We have verified that the approximation above reproduces the full Sommerfeld calculation to

[5] Additional constraints can be derived from neutrino experiments. These constraints are typically weaker than those arising from gamma ray and antiproton data (see e.g. [105] for discussion of the ν flux in the context of a related model) and we do not consider them in this work. Under specific cosmic ray propagation model assumptions, constraints can also be derived from the high energy positron flux. In comparison to the $\bar{p}$ calculation, however, the theoretical uncertainties for e^+ are larger as the results depend crucially on the cosmic ray propagation time in the Galaxy that dictates the amount of e^+ radiative energy loss [114], and so we do not consider e^+ constraints in this work.

a good accuracy.

The top panel of Fig. 3.4 shows that for DM mass above a few TeV, large values of the SE factor are possible with $SE_{eff} > 10^2$ in resonance regions. As we show below, this result may have interesting implications – striking indirect detection signatures are possible if the model happens to live at an SE resonance. However, resonant SE is limited to fine-tuned regions in the parameter space. To illustrate this point, in the bottom panel we plot the value of SE vs. the DM mass fixing $m_\sigma = 3$ TeV (corresponding to a vertical slice through the center of the top panel, marked by an arrow). For generic parameter configuration the effective SE factor is modest, and only grows above 10^2 near resonances and for extremely heavy DM mass, close to the unitarity limit where our calculation breaks down. Note that we truncate the value of SE_{eff} at 10^3 in resonance peaks. As the resonance regions are fine-tuned, this has limited impact on our analysis. According to the analysis of [87], the relic abundance is depleted at the tip of these SE resonances due to chemical re-coupling of DM at low redshifts, an effect that we do not include here and that would reduce the value of SE_{eff}. In addition, the low velocity divergence of the SE at the resonance tip should be regulated by bound-state decay that would also suppress the peak SE.

In Secs. 3.6.1 and 3.6.2 below we calculate antiproton and gamma ray constraints on the model. For antiprotons we adopt a conservative model-independent approach to the problem of cosmic ray propagation, and provide some extra details to explain our method. The summary of our results is that the bulk of the parameter space of Fig. 3.4 (or equivalently Fig. 3.2) is allowed by current constraints. This is not a surprise: much of the parameter space consistent with the DM relic density corresponds to rather heavy m_χ at the several TeV, where current indirect searches do not yet constrain the thermal relic cross section. Indirect detection constraints do exclude, or make promising predictions for, the near-resonant SE regions seen in Fig. 3.4. If one accepts the assumption of a cusp DM

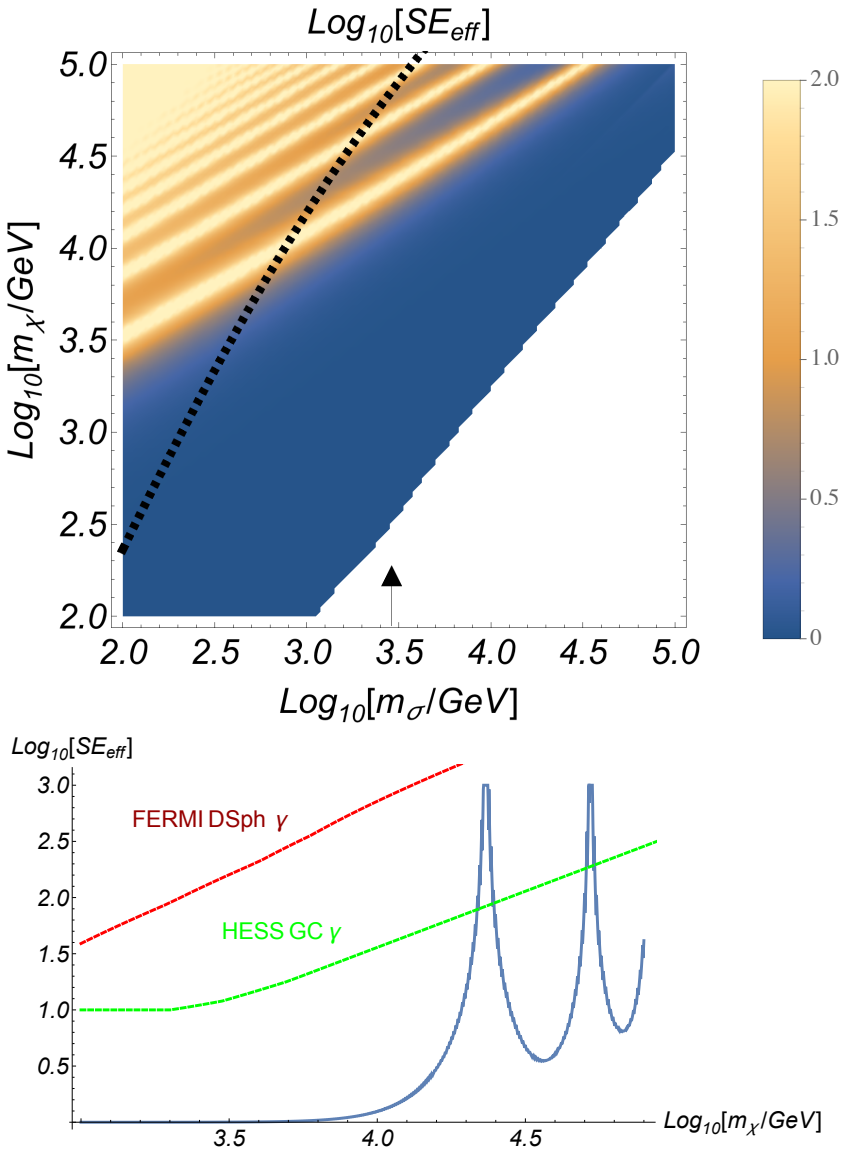

Figure 3.4: Top panel: Sommerfeld enhancement factor (SE) in the $\{m_\sigma, m_\chi\}$ plane. Above the dashed line $m_\sigma < f/10$, indicating fine-tuning. Bottom panel: SE vs. dark matter mass, fixing the dilaton mass to $m_\sigma = 3$ TeV (marked on top panel with an arrow). The region above the red and green dashed lines is excluded by FERMI and HESS gamma ray observations (the latter depend strongly on assumptions regarding the DM distribution in the Galaxy; see Sec. 3.6.2). The dark matter particles' relative velocity today is set to $v = 10^{-3}$.

density profile in the Milky Way Galactic Center, for example, then HESS gamma ray data already excludes much of the parameter region in the upper-left corner of the top panel of Fig. 3.4. Finally in 3.6.3 we end with a brief discussion of the constraints on dark matter self-interaction.

3.6.1 Antiprotons

The PAMELA satellite experiment reported a measurement of the high energy antiproton flux in interstellar space, extending up to 350 GeV [117]. The PAMELA measurement is consistent with model-independent calculations of the antiproton flux expected due to fragmentation of high energy primary cosmic ray nuclei on ambient interstellar gas in the Galaxy [118].

Following Ref. [105], we derive a bound on the antiproton production in dark matter annihilation by imposing that the dark matter annihilation source of antiprotons in the local Galactic gas disc does not exceed the source due to the astrophysical production, in the energy range covered by the current measurements. The bound derived in this manner is independent of modeling assumptions regarding the propagation of charged cosmic rays in the Galaxy. The bound is conservative because it does not include the possible additional contribution of DM annihilation in the cosmic ray halo that may extend well above and below the gas disc.

The injection rate density of antiprotons due to DM annihilation is given by

$$\begin{aligned} Q_{\bar{p},DM}(E) &= \frac{1}{2} n_\chi^2 \langle \sigma v \rangle \frac{dN_{\bar{p}}}{dE} \approx 5 \times 10^{-36} \mathrm{cm}^{-3} \mathrm{s}^{-1} \mathrm{GeV}^{-1} \times \\ &\left(\frac{\rho_\chi}{0.4\ \mathrm{GeV cm}^{-3}} \right)^2 \left(\frac{\langle \sigma v \rangle}{3 \times 10^{-26}\ \mathrm{cm}^3 \mathrm{s}^{-1}} \right) \left(\frac{m_\chi}{1\ \mathrm{TeV}} \right)^{-3} \left(m_\chi \frac{dN_{\bar{p}}}{dE} \right). \end{aligned} \quad (3.28)$$

Here, $\rho_\chi = m_\chi n_\chi \approx 0.4\ \mathrm{GeV\,cm}^{-3}$ is the DM mass density in the local halo and $\frac{dN_{\bar{p}}}{dE}$ is the differential antiproton spectrum per annihilation event. To compute $\frac{dN_{\bar{p}}}{dE}$ we use the code

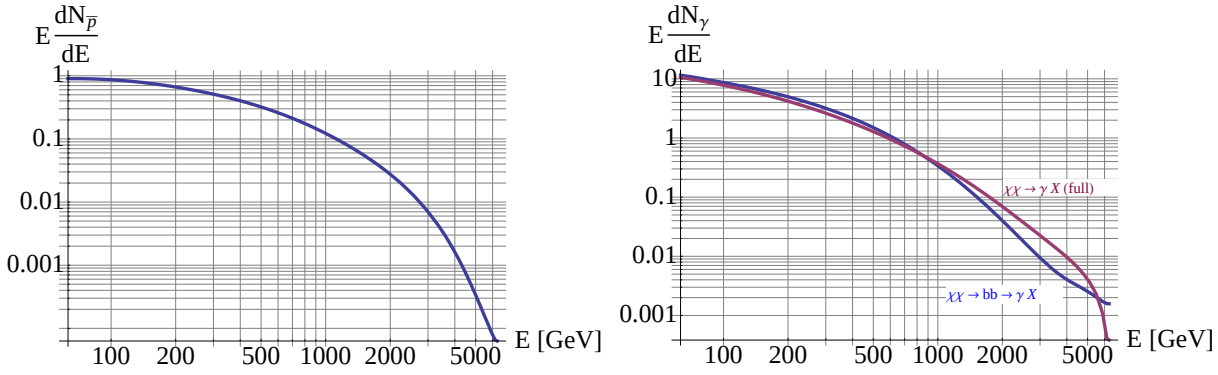

Figure 3.5: Left: differential $\bar{p}$ spectrum per DM annihilation, computed for $m_\chi = 6.3$ TeV, $m_\sigma = 427$ GeV, and $f = 6.2$ TeV. Right: same for the gamma ray spectrum; the purple line shows the spectrum from the full annihilation process including all dominant partial channels ($\chi\chi \to WW, ZZ, tt, \sigma\sigma, ...$), while the blue line shows the spectrum due to $\chi\chi \to bb$ alone.

provided in Ref. [119], that directly produces the differential $\bar{p}$ spectrum for the channels $\chi\chi \to WW, ZZ, hh, t\bar{t}$ accounting for the decay and hadronization of the intermediate unstable states. To include the contribution of $\chi\chi \to \sigma\sigma$ we proceed in two steps. First we use Ref. [119] to calculate the $\bar{p}$ spectrum arising in the dilaton rest frame due to dilaton decay; define this spectrum by $\left[\frac{dN_{\bar{p}}}{dE}(E)\right]_{\sigma\to\bar{p}X}$. We then convolve the dilaton decay $\bar{p}$ spectrum with the isotropic boost factor of the σ in the DM annihilation center of mass frame, obtaining

$$\left[\frac{dN_{\bar{p}}}{dE}(E)\right]_{\chi\chi\to\sigma\sigma} = \frac{1}{\gamma_\sigma\beta_\sigma}\int_{\beta_\sigma^{-1}-1}^{\beta_\sigma^{-1}+1}\frac{dx}{x}\left[\frac{dN_{\bar{p}}}{dE}\left(\frac{E}{x\gamma_\sigma\beta_\sigma}\right)\right]_{\sigma\to\bar{p}X} \tag{3.29}$$

where $\gamma_\sigma = m_\chi/m_\sigma$ and $\beta_\sigma = \sqrt{1-\gamma_\sigma^{-2}}$. We neglect DM annihilation into gluons, since the branching fraction of annihilation to this state is small compared to that of annihilation to quarks and massive gauge bosons. In the left panel of Fig. 3.5 we plot the differential flux of $\bar{p}$ from DM annihilation with $m_\chi = 6.3$ TeV, $m_\sigma = 427$ GeV, and $f = 6.2$ TeV reproducing the observed DM relic abundance.

The injection rate density due to primary cosmic rays colliding with interstellar gas in

the disc is [118]

$$Q_{\bar{p},CR}(E) \approx 8.4\times10^{-33}\mathrm{cm}^{-3}\mathrm{s}^{-1}\mathrm{GeV}^{-1}\times \left(\frac{E}{100\ \mathrm{GeV}}\right)^{-2.8}\left[1-0.22\log_{10}^2\left(\frac{E}{500\ \mathrm{GeV}}\right)\right]\frac{J_p(1\ \mathrm{TeV})}{J_{p,0}(1\ \mathrm{TeV})}, \quad (3.30)$$

where $J_p(1\ \mathrm{TeV})$ is the local proton flux sampled at $E = 1\ \mathrm{TeV}$ and scaled to the measured value $J_{p,0}(1\ \mathrm{TeV}) \approx 8\times10^{-9}\ \mathrm{GeV}^{-1}\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$. The uncertainties in the derivation of Eq. (3.30) are at the ∼50% level. Our conservative bound on the DM annihilation rate amounts to imposing that the ratio $Q_{\bar{p},CR}(E)/Q_{\bar{p},DM}(E)$ is larger than unity for E in the range 10-300 GeV.

The basic result we find is that the model survives our antiproton constraint by a large margin, unless it lives right on top of an SE resonance. If the model is near an SE resonance, then a detectable rise in the antiproton flux at high energy is predicted. For DM mass below about ~ 10 TeV, the rise would be in tension with currently available $\bar{p}$ data and the model is observationally disfavored (again, only the region near an SE resonance, as seen in Fig. 3.4). For $m_\chi \gtrsim 10$ TeV, though, the rise in the $\bar{p}$ flux sets in at high energy with only a moderate effect in the energy range where current data exists. In this case, improved high energy cosmic ray measurements expected in the near future [120] may detect the model in the $\bar{p}$ flux.

We illustrate these findings in Fig. 3.6 where we plot the expected antiproton flux in our model near an SE resonance for two chosen points. The data points (last one being an upper bound) and the green curve denote the PAMELA data and the secondary astrophysics prediction, respectively. The red and magenta curves give an estimate of the antiproton flux that would occur for the parameter points $\{m_\chi = 6.3\ \mathrm{TeV}, m_\sigma = 300\ \mathrm{GeV}\}$ and $\{m_\chi = 31\ \mathrm{TeV}, m_\sigma = 4.7\ \mathrm{TeV}\}$, respectively, where the effective SE factor is $SE_{eff} \approx 10^3$ (fixing f to obtain the observed DM relic abundance).

Above we chose tuned points with large SE_{eff} to illustrate the possible $\bar{p}$ signal; as

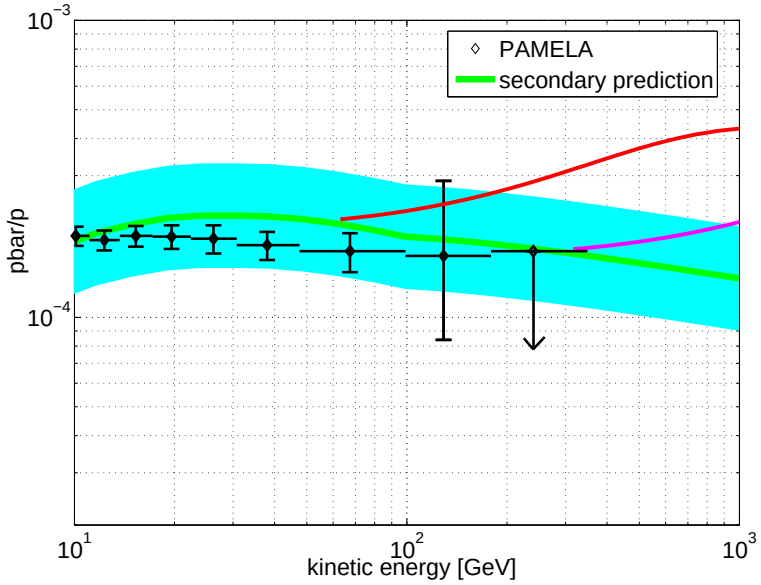

Figure 3.6: Antiproton flux with DM annihilation at a Sommerfeld factor resonance. Data points and green curve denote PAMELA data and secondary astrophysics prediction, respectively. Red and magenta curves give a lower estimate of the $\bar{p}$ flux with DM annihilation for the model parameter point with $\{m_\chi = 6.3\ \mathrm{TeV}, m_\sigma = 300\ \mathrm{GeV}\}$ and $\{m_\chi = 31\ \mathrm{TeV}, m_\sigma = 4.7\ \mathrm{TeV}\}$, respectively, where the SE factor is $SE_{eff} \approx 10^3$.

mentioned earlier, this large SE near the resonance peak can be damped somewhat by a more careful treatment of the relic abundance. However, we stress that the DM-induced signal depends on unknown cosmic ray propagation features. The red and magenta curves in Fig. 3.6 should be considered as a robust *lower bound* on the DM-induced flux. Considering disc+halo diffusion models [121], for example, the actual flux could be as high as a factor of ~ 100 above the result we show here[6]. A future detection of the model through cosmic ray $\bar{p}$ is therefore conceivable also away from SE resonance peaks.

[6]See App. B of Ref. [105] for a detailed discussion.

3.6.2 Gamma Rays

The FERMI gamma ray telescope reported limits on DM annihilation based on a stacking analysis of dwarf spheroidal galaxies [122]. The analysis is relatively insensitive to the assumed DM mass distribution in the target galaxies. Ref. [122] reports limits directly on the annihilation cross section for the specific channel $\chi\chi \to b\bar{b}$ as a function of the DM mass. Using the code of Ref. [119] and following a similar method as that described above for the $\bar{p}$ spectrum calculation, we verified that the spectrum of continuum gamma rays obtained in our model agrees to within a factor of 2-3 with the gamma ray spectrum resulting from a pure $\chi\chi \to b\bar{b}$ channel. In what follows we therefore assume that the constraints quoted in [122] apply to our model directly. In the right panel of Fig. 3.5 we plot the differential gamma ray flux from DM annihilation with $m_\chi = 6.3$ TeV, $m_\sigma = 427$ GeV, and $f = 6.2$ TeV reproducing the observed DM relic abundance. The purple line shows the spectrum from the full annihilation process including all dominant partial channels ($\chi\chi \to WW, ZZ, tt, \sigma\sigma, ...$), while the blue line shows the spectrum due to $\chi\chi \to bb$ alone.

We extrapolate the bound to $m_\chi = 100$ TeV, using the scaling $m_\chi^{-2}\frac{dN_\gamma}{dE} \sim m_\chi^{-1}$ that applies for photon energies in the FERMI range, $E \lesssim 500$ GeV $\ll m_\chi$. The resulting bound is illustrated by the red dashed line in the bottom panel of Fig. 3.4, focusing on a slice in the parameter space with $m_\sigma = 3$ TeV.

Stronger, but more model-dependent limits are obtained from ground-based air-Cherenkov telescopes. The HESS gamma ray observatory reported limits on DM annihilation based on Galactic Center observations [123]. Due to the background subtraction method of the experiment, the analysis is not sensitive to shallow DM density profiles, and so the results are only applicable under the assumption of a cusp profile such as the Navaro-Frenk-White [124] distribution. Assuming a cusp distribution, neglecting the $\mathcal{O}(1)$ spec-

tral difference between the $\chi\chi \to q\bar{q}$-induced gamma ray spectrum assumed in [123] and the actual spectrum in our model, and extrapolating their limits from $m_\chi = 10$ TeV up to $m_\chi = 100$ TeV, we obtain the bound depicted by the green dashed line in the bottom panel of Fig. 3.4.

Finally, both FERMI [37] and HESS [125] reported limits on DM annihilation to a gamma ray line. We calculate the branching fraction $\langle\sigma v\rangle(\chi\chi \to \gamma\gamma)/\langle\sigma v\rangle(\text{total})$ using Eq. (E.3). This branching fraction is very small in our model, reminiscent of the result for a heavy Higgs. Consequently the gamma ray line constraint is sub-dominant compared to the continuum emission bounds. We comment that the HESS limit [125] have recently been used to put significant pressure on supersymmetric Wino dark matter [126, 127]. This situation is not reproduced here; for the Wino example, the strong exclusion is primarily due to the presence of an electromagnetically charged state that is mass-degenerate with the neutral DM particle, amplifying the di-photon annihilation diagram. Without a special construction of this kind, our dilaton-mediated DM scenario passes the line searches unscathed.

3.6.3 Self-interaction and dipoles

Self-interaction of dark matter could occur in our model via two mechanisms. First, dilaton exchange leads to self-interaction cross sections per unit mass that scale roughly as $\sigma/m_\chi \sim m_\chi/f^4$ (assuming $m_\sigma \lesssim m_\chi$; otherwise there is an additional suppression by $(m_\chi/m_\sigma)^4$). Imposing the unitarity constraint as implemented in Eq. (3.18), $m_\chi \lesssim 5f$, we have $\sigma/m_\chi < 600/m_\chi^3 \sim 10^{-10}\,(\text{TeV}/m_\chi)^3$ cm^2/g. Throughout our relevant parameter space, this is many orders of magnitude below the observational constraints on the self interaction of dark matter, of order ~ 1 cm^2/g [17]. A second potential source of dark matter self-interaction could arise from the CFT side. Given the weakness of the astro-

physical constraints, some amount of self-interaction is not excluded. However, without accounting for the details of the assumed (broken) CFT in which χ is embedded, we cannot address this issue quantitatively.

Similarly, limits from electric or magnetic moments are not constraining this model. Indeed, by assumption, our dark matter is only connected to the SM fields via dilaton exchange. Thus any magnetic or electric dipole moment must be induced through dilaton exchange (the latter necessitating, in addition, CP violation), with the exchanged particle being at the TeV scale, leading to induced couplings that are much too small to affect cosmological or astrophysical observations. For instance, considering the magnetic dipole moment (MDM) limit reported in [128], an MDM of $\mathcal{M} \sim 1/\mathrm{GeV}$ is still easily consistent with cosmological data.

3.7 Discussion

In this paper we explored the possibility that the dilaton could mediate dark matter annihilation. Such models have the appeal that the couplings are largely determined by scale invariance. The breaking scale of scale invariance f is fixed by requiring that the relic abundance matches the observed value, leaving the dark matter and dilaton masses as the main free parameters. We mapped the relevant $\{f, m_\chi, m_\sigma\}$ parameter space taking the various dark matter annihilation modes into account and imposing unitarity bounds. We showed that large regions of parameter space, with f, m_χ, m_σ all in the $\sim 1 - 10$ TeV range, can correctly reproduce the observed relic abundance. We find an upper bound $f \leq 30 - 100$ TeV, implying a similar bound on $m_{\sigma,\chi}$.

Collider searches for Higgs-like particles, including LHC, Tevatron and LEP analyses, put model dependent lower bounds on f for dilaton masses up to ~ 1 TeV. The collider

bounds exclude dilaton-mediated dark matter for $m_\chi \lesssim 200$ GeV. Current direct detection experiments yield similar model dependent exclusions for the lower end of the mass spectrum, requiring $m_\chi \gtrsim 300$ GeV for $m_\sigma \lesssim 300$ GeV. The predicted dark matter-nucleon elastic scattering cross section becomes independent of the dark matter mass for heavy dark matter.

Our analysis of indirect detection included antiproton and gamma ray data and shows that the bulk of the parameter space is consistent with the current constraints. A possible signal in high energy cosmic ray antiprotons could appear for favorable cosmic ray propagation scenario for models with parameters close to a Sommerfeld enhancement resonance. A promising avenue for probing the model all the way to very high DM mass is in high energy ground-based gamma ray measurements, see e.g. [129, 130] for recent reviews. For scalar or vector DM, future gamma ray experiments should detect or exclude the entire parameter space of the model.

CHAPTER 4

PLAYING TAG WITH ANN: BOOSTED TOP IDENTIFICATION WITH PATTERN RECOGNITION

When high-momentum heavy particles are produced at the LHC their decay products may form overlapping jets in the detector. At first glance, these jets look very similar to the regular QCD generated jets. Nevertheless, in recent years several algorithms have been developed to distinguish the two kinds of jets. In this chapter we develop our own algorithm which uses a popular image classification method, namely artificial neural networks.

This chapter is based on the paper *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition* [131], written in collaboration with Leandro G. Almeida, Mihailo Backovic, Seung J. Lee and Maxim Perelstein.

4.1 Overview

Many searches for physics beyond the Standard Model at the Large Hadron Collider (LHC) rely on top tagging algorithms, which discriminate between boosted hadronic top quarks and the much more common jets initiated by light quarks and gluons. We note that the hadronic calorimeter (HCAL) effectively takes a "digital image" of each jet, with pixel intensities given by energy deposits in individual HCAL cells. Viewed in this way, top tagging becomes a canonical pattern recognition problem. With this motivation, we present a novel top tagging algorithm based on an Artificial Neural Network (ANN), one of the most popular approaches to pattern recognition. The ANN is trained on a large sample of boosted tops and light quark/gluon jets, and is then applied to independent test samples. The ANN tagger demonstrated excellent performance in a Monte Carlo study: for example, for jets with p_T in the $1100 - 1200$ GeV range, 60% top-tag efficiency can be achieved with a 4% mis-tag rate. We discuss the physical features of the jets identified by the ANN tagger as the most important for classification, as well as correlations between the ANN tagger and some of the familiar top-tagging observables and algorithms.

4.2 Introduction

Many extensions of the Standard Model (SM) predict new particles with masses around the TeV scale. Searches for such new particles form a major component of the experimental program at the Large Hadron Collider (LHC). In most models, the new particles are unstable, and their decays often contain weak-scale SM states, namely the W and Z bosons, the Higgs boson, and the top quark. Searches for final states containing top quarks are particularly important, due to the special role played by the top sector in many models of electroweak symmetry breaking. Decays of heavy new particles with mass

above the electroweak scale typically result in highly energetic, relativistic top quarks in the lab frame. Identifying and characterizing such "boosted" top quarks in the data is crucial for new physics searches and tests of naturalness [132] at the LHC, especially as the bounds on the new physics mass scales in many candidate models are pushed higher. Examples of new physics leading to boosted top signatures include Kaluza-Klein gluons [133, 134] and string Regge states [135] of the Randall-Sundrum model, stops [136] and gluinos [137] of supersymmetry, top and light quark partner decays in Composite Higgs models [138, 139, 140, 141, 142, 143], and many others.

Due to relativistic kinematics, the decay products of a boosted top quark are highly collimated. For instance, hadronic decay of a top quark of $p_T \sim 1$ TeV would produce three quarks collimated into a cone of rough size $R \sim 0.4$ and result in a specific pattern of hadronic activity in the detector. Classical event reconstruction techniques are inadequate to tag and measure such topologies, as most of the showered radiation falls into a small angular region. One solution is to cluster the event with a large jet cone ($R \sim 1$), and consider the features of energy distribution inside such "fat" jets (so-called jet substructure), instead of correlations between individual small radius jets. Over the past decade, a variety of methods for boosted top tagging via jet substructure have been developed (see Ref. [144] for a review), most of which can be cast into several (non exclusive) groups. Jet shapes are observables based on various moments of the jet energy distribution. Notable examples are angular correlations studied extensively in Ref. [145], sphericity tensors [146, 147] and other perturbatively calculable jet shapes [148]. Considerations of jet clustering history led to development of numerous Filtering jet substructure methods [149, 150, 151], where the differences in the late steps of jet clustering between heavy SM states and QCD jets from light partons have been successfully applied in tagging of heavy SM states. Furthermore, Prong Taggers such as N-subjettiness [152, 27] exploit the differences in the number of hard energy depositions within the boosted jet (e.g. three-body top decays compared to the typical two-body splitting of a light jet). Parton level

models of boosted decays and kinematic constraints built into them can also be used to study jet substructure, with the Template Overlap Method (TOM) [153, 154, 155, 156] being the most notable example. More recently, Matrix Element Method [157, 158] inspired techniques such as Shower Deconstruction have emerged [159, 160], where a boosted jet is tagged using approximations to hard matrix elements and the parton shower. Soft drop declustering (a generalization of modified mass drop tagging) is another method which has been recently developed for removing non-global contributions (soft radiation) to the jet [161]. Several of these methods have been implemented in the analyses of the LHC data by the CMS and ATLAS collaborations; see, for example Ref. [162, 163, 164, 165].

In this paper, we pursue an alternative approach to jet substructure. Experimentally, information about hadronic activity in an event comes mainly from the hadronic calorimeter (HCAL), with the basic observable being the energy deposited in each of the HCAL cells. One can think of the information provided by the HCAL as a *digital image*, with each cell (or topo-cluster) being identified as a pixel, and with energy deposit in the cell corresponding to the intensity (or grayscale color) of that pixel. From this point of view, boosted top identification is simply a classic image-recognition problem: distinguishing the energy-deposit patterns characteristic of boosted tops from patterns due to other sources, such as the usual QCD jets. This suggests that computational algorithms developed in the field of image recognition could be of use in boosted top tagging.[1]

With this motivation, we constructed a new top tagger algorithm based on one of the most popular approaches to image recognition, Artificial Neural Networks (ANNs). In this approach, each jet is classified as top or non-top according to a highly non-linear scoring function. The function contains multiple adjustable parameters, called weights. These

[1]Recently, Ref. [28] studied jet substructure as an image recognition problem in the context of boosted W tagging as well gluon/quark discrimination. The authors utilised a linear Fisher discriminant trained on a sample of signal and background events, in order to distinguish the desired events from the backgrounds. The method out-performs the existing methods of W tagging, illustrating the benefits of the image recognition approach to jet substructure.

are chosen using a training procedure, in which the ANN is presented with a large sample of jets that are known to be top or non-top, and the weights are chosen to maximize the number of correctly identified jets in this sample. (In our study, all samples are generated by Monte Carlo simulations. In experimental applications, ANN may be trained on either MC samples or carefully selected "calibration" data sets.) Having fixed the weights, the ANN is then applied to independent samples containing both top and non-top jets, and asked to discriminate between them. We find that the performance of the ANN tagger significantly exceeds that of several popular tagging algorithms currently in use over a wide range of p_T, demonstrating the practical utility of this approach.

The paper is organized as follows. Section 4.3 describes the MC event samples used for training and testing the ANN tagger, as well as the pre-processing steps applied to these samples before the ANN is applied. Section 4.4 contains a detailed description of the ANN tagger, including the network architecture and the training algorithms we employed. In Section 4.5, we present the results of our study of ANN tagger performance and comparisons with other popular taggers. We also discuss the physical features of jets that are dominant in the ANN classification, and the extent to which ANN output is correlated with that of other taggers. We conclude with a recap and a brief discussion of directions for future research in Section 4.6. An Appendix contains a brief description of the top taggers we use for the purpose of comparison with the ANN tagger.

4.3 Event Generation and Pre-Processing

We generate benchmark event samples with `MadGraph 5` [166] at leading order, and shower them with `Pythia 6` [167]. In order to study the effects of different showering algorithms on the results, we also generate separate data samples showered with `Pythia 8` [168]. For simplicity, we extract a pure sample of top jets from a Standard

Model top pair-production simulation, at leading order with no matching. The tops are decayed in `MadGraph 5`, so that the angular distribution of the decay products is modeled correctly. Similarly, we generate the light jet sample from a simulation of the QCD di-jet process, including both quarks and gluons in the final state, but no matching to extra jets. Fiducial cut $|\eta| \leq 5.0$ is imposed at the hadron level. We cluster the events using the `fastjet` [169] implementation of the anti-k_T algorithm [170] with a large jet cone of $R = 1.0$. For our analysis, we only use the highest p_T jet in each event, and impose the cut $|\eta_{\text{jet}}| \leq 2.5$. We consider samples of jets within three jet p_T ranges: $500 - 600$ GeV, $800 - 900$ GeV and $1100 - 1200$ GeV. These three bins span a range of jet p_T values relevant for top tagging at the LHC, while analyzing them separately provides information about p_T sensitivity of the tagging efficiency and other parameters. Unless otherwise noted, we impose a cut on the jet mass (*i.e.* the invariant mass of all particles assigned to the jet), selecting jets within a window

$$130 \text{ GeV} < m_J^{R=1.0} < 210 \text{ GeV}. \tag{4.1}$$

A vast majority of top jets fall within this mass range, while most QCD jets are rejected by this cut. Discriminating the remaining QCD jets from top jets is the task for the top tagger.

In order to form an input to the ANN tagger, we preprocess each jet as follows. First, we find the center of the jet, defined by the sum of the coordinates of all particles weighted by their energies,

$$\eta_C = \frac{1}{E}\sum_j \eta_j E_j, \;\; \phi_C = \frac{1}{E}\sum_j \phi_j E_j, \tag{4.2}$$

where $E = \sum_j E_j$ is the total energy of the jet. We then shift the coordinates of each particle so that the jet is centered at the origin in the new coordinates:

$$\eta'_j = \eta_j - \eta_C, \;\; \phi'_j = \phi^j - \phi_C. \tag{4.3}$$

Further, we find the jet "principal axis" in the (η, ϕ) plane, defined by

$$\tan(\theta) = \frac{\sum_j \frac{\phi'_j \cdot E_j}{\Delta R'}}{\sum_j \frac{\eta'_j \cdot E_j}{\Delta R'}}, \quad \Delta R' = \sqrt{\eta'^2_j + \phi'^2_j}, \tag{4.4}$$

and rotate the coordinate system so that this principal axis is the same direction $(+\eta)$ for all jets:

$$\eta''_j = \eta'_j \cdot \cos(\theta) + \phi'_j \cdot \sin(\theta), \tag{4.5}$$

$$\phi''_j = -\eta'_j \cdot \sin(\theta) + \phi'_j \cdot \cos(\theta). \tag{4.6}$$

These coordinate transformations remove information about the jet position in the calorimeter and its orientation in the (η, ϕ) plane. Both pieces of information are irrelevant for top tagging, and removing them from consideration allows the ANN tagger to focus on the irreducible physical differences between top and QCD jets.[2]

In the new coordinates, nearly all (98%) of the particles assigned to a given jet fall within a window of $\eta'' \in [-\pi/2, \pi/2]$ and $\phi'' \in [\pi/2, \pi/2]$. We model the HCAL response to the jet by dividing this window into 30×30 square cells. (The cell size is approximately 0.1×0.1, close to the realistic values in ATLAS and CMS.) The normalized energy deposited in each cell, ε_{ab} $(a, b = 1 \ldots 30)$, is computed by adding up the energies of all particles falling within that cell, and dividing by the total energy of the jet. (The last step is once again necessary to remove information irrelevant for top tagging, in this case the total jet energy.) By construction, ε_{ab} is dimensionless and lies between 0 and 1. In the language of image processing, each jet has been converted into an image with 30×30 pixels, with a grayscale color of each pixel given by the corresponding ε_{ab}. These images can now be classified by an Artificial Neural Network (ANN), described in the following section.

[2] As an exercise, we also attempted to train the neural network on a set of jets with randomly oriented principal axes, *i.e.* without the rotation (4.6). We found that this procedure still yields an effective tagger; presumably, the neural net learns to ignore the axis orientation information during the training process. However, to achieve the same tagging performance, the randomly-oriented training set needs to be significantly larger.

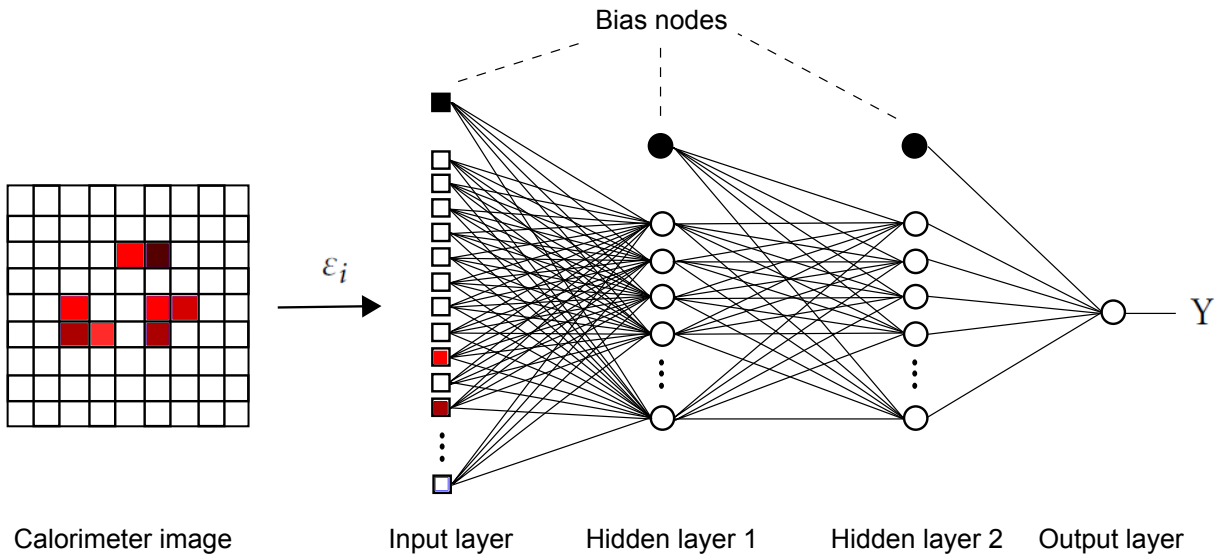

Figure 4.1: Graphical representation of the Artificial Neural Network (ANN).

4.4 ANN Tagger

ANN tagger is based on a feed-forward neural network with an input layer consisting of $30 \times 30 = 900$ nodes, one for each calorimeter cell; two hidden layers, of 100 nodes each, to process the signal; and an output layer consisting of a single node, whose value Y is interpreted as the probability that a given jet comes from a boosted top decay. The inputs to the neural network ϵ_i are simply the normalized energy deposits ε_{ab} defined above, rearranged in a single 900-dimensional vector: $\varepsilon_{ab} \equiv \epsilon_{30a+b}$. The architecture of the network is shown in Fig. 4.1. (For a pedagogical introduction to Artificial Neural Networks in the context of image recognition, see Section 1.2.2.)

To train the network, we use a set of $N/2$ top and $N/2$ QCD jets, where N is a large number. For the i-th jet, we assign the "target output" variable: $y_i = 1$ if it is a top jet, and $y_i = 0$ if it is a QCD jet. Training consists of adjusting the weights so that the actual outputs of the ANN Y_i correspond as close as possible to the target outputs y_i, across

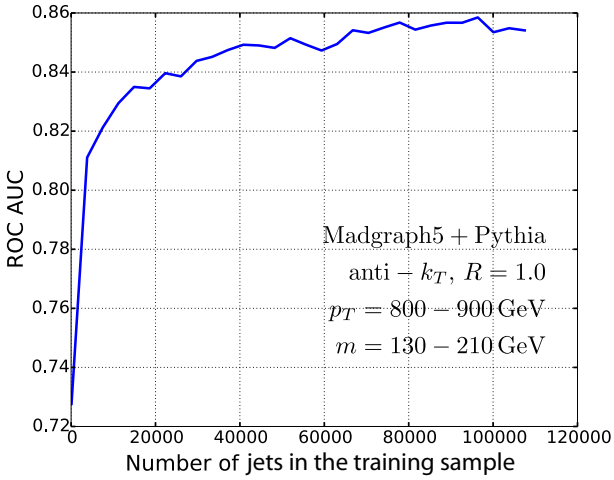

Figure 4.2: ROC AUC on a cross-validation set of 50 000 jets, vs. number of jets in the training set.

the training set. To quantify the error, we use the logarithmic loss variable and to train the neural network we use the back-propagation algorithm [24] combined with gradient-descent minimization.

Over several experiments we found that the easiest way to avoid over-fitting was to use more training data and ensembling several neural networks together. To determine the size of the training set N_{tr} needed to saturate the learning of our neural network, we studied the performance of the trained network on a cross-validation set of 50000 top and QCD jets, as a function of N_{tr}. For this analysis, the performance is characterized by the ROC AUC (area under the receiver operating characteristic curve) performance metric, which assigns a value of 0.5 to a random classifier and a value of 1.0 to a perfect classifier. As can be seen on Fig. 4.2, performance steadily improves with the training set size until $N_{\mathrm{tr}} \approx 40000$ (*i.e.* 20000 top images and 20000 dijet images), after which convergence is achieved. This indicates minimal over-fitting beyond that point.

To further improve the performance of our tagger, we ensembled multiple neural networks together. The idea is to train B neural networks together, with the output given by

the average of their outputs,

$$\mathcal{O} = \frac{1}{B}\sum_{i=1}^{B} Y_i. \tag{4.7}$$

In our application, $B = 10$. All networks are trained using the same training set, but the jets are weighted. For the first network, all weights are set to one. Jets which are heavily misclassified by the first network are then assigned a larger weight, while jets which are correctly classified are assigned a smaller weight. This re-weighted training set is then used to train the second network, and so on. This procedure allows the training algorithm to focus on specific events that are particularly arduous to classify, improving overall performance. For some parameter choices, this method can be mapped to boosted methods such as ADAboost [171], where the weak classifiers are feed-forward ANNs.

4.5 Results

The ensemble of ANNs described above has been trained on sets of about 50,000 top and QCD jets each, in three p_T bins, $500 - 600$ GeV, $800 - 900$ GeV, and $1100 - 1200$ GeV. It has then been applied to test sets consisting of about 15,000 top and QCD jets each, in the same p_T bins. The distribution of the neural network output $\mathcal{O}$ on the test sets is shown in Fig. 4.3. The classification power of this observable is clear from the figure: top jets are predominantly assigned $\mathcal{O} \approx 1.0$, while QCD jets are predominantly assigned $\mathcal{O} \approx 0.0$. To use the ANN ensemble as a top-tagger, we simply choose a threshold value $\mathcal{O}_{\rm th}$, and assign the "top tag" to any jet with $\mathcal{O} \geq \mathcal{O}_{\rm th}$ and the "QCD tag" to any jet with $\mathcal{O} < \mathcal{O}_{\rm th}$.

To discuss the performance of the ANN tagger, it is convenient to define efficiency and mis-tag rates as follows:

$$\text{Eff} = \frac{N^{\rm top}_{\rm top}}{\mathcal{N}_{\rm top}}, \quad \text{Mistag} = \frac{N^{\rm top}_{\rm QCD}}{\mathcal{N}_{\rm QCD}}, \tag{4.8}$$

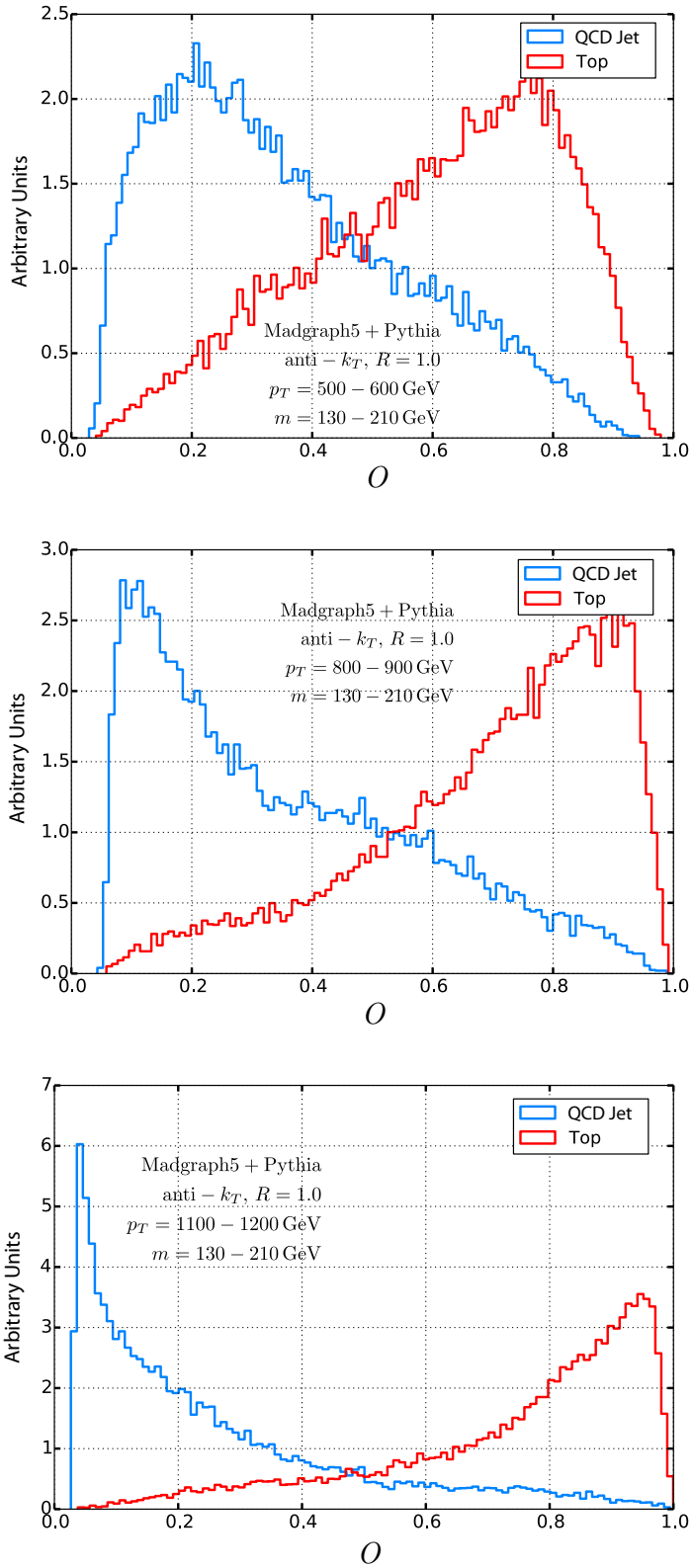

Figure 4.3: Distributions of the ANN output $\mathcal{O}$ on top (red) and QCD (blue) jet samples in three representative p_T ranges. All distributions are normalized to unit area.

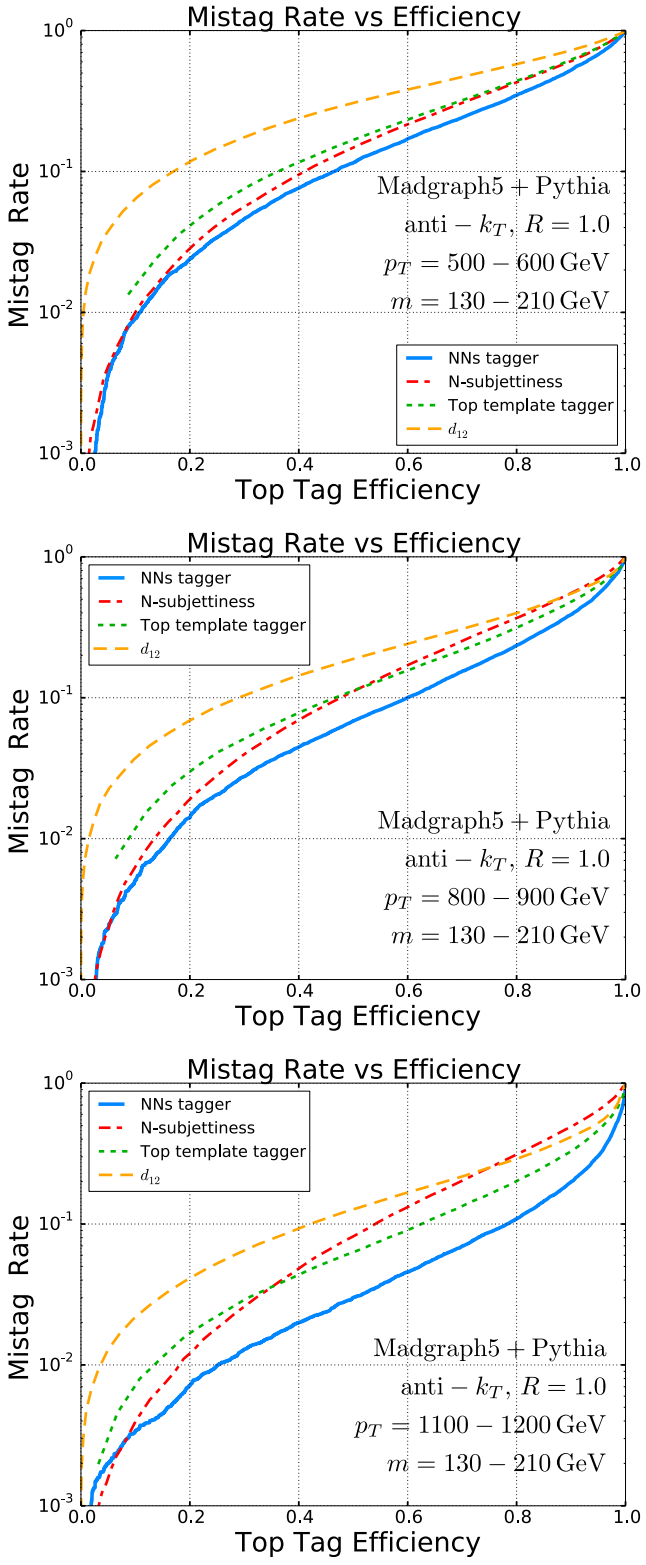

Figure 4.4: Efficiency vs. Mis-tag rate curves for the ANN tagger (blue/solid lines), for jets in three representative p_T ranges. For comparison, corresponding curves for three existing top taggers are also shown: d_{12} tagger (yellow/dashed), top template tagger (green/dotted), and N-subjettiness (red/dash-dotted).

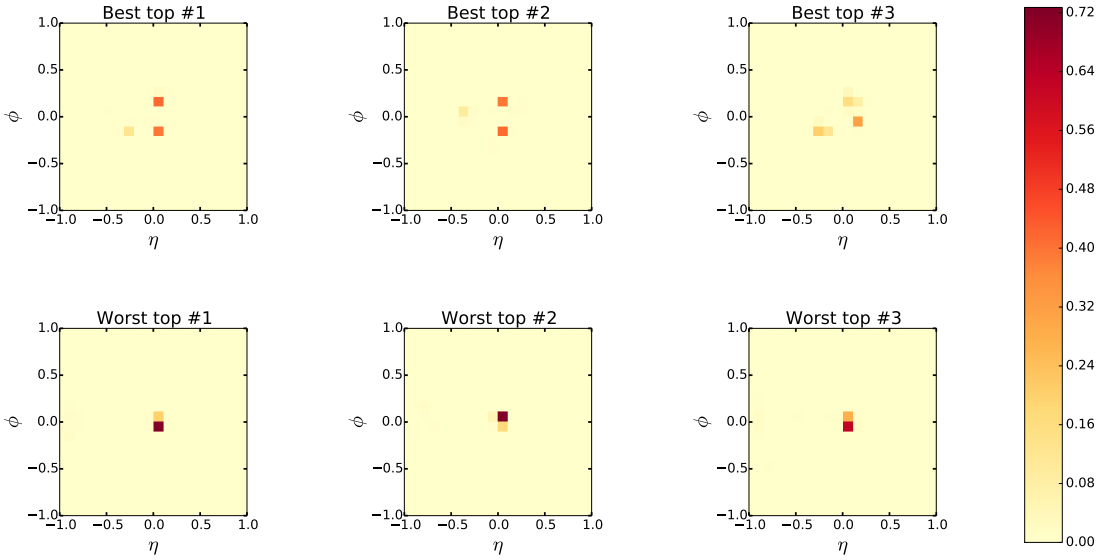

Figure 4.5: Energy deposit patterns for three jets with the highest (top row) and lowest (bottom row) ANN scores in the top sample with $p_T \in [800, 900]$ GeV.

where $\mathcal{N}_{\rm top}$ and $\mathcal{N}_{\rm QCD}$ are the total number of jets in the top and QCD jet samples, respectively, and N_a^b is the number of jets in sample a tagged as jets of type b (a, b =top, QCD). Efficiency and mis-tag rates can be varied by varying the threshold $\mathcal{O}_{\rm th}$. The performance of the ANN tagger is shown in Fig. 4.4, where for comparison we also show the performance of three representative existing taggers, described in the Appendix. In all cases, the ANN tagger outperforms the existing taggers, achieving lower mis-tag rates for the same tagging efficiency. The improvement is especially dramatic for high jet p_T: for example, for jets with $p_T \in [1.1, 1.2]$ TeV range, the ANN tagger achieves 60% tagging efficiency with about 4% mis-tag rate, about a factor of 2 lower than the best of the existing taggers in our comparison pool. This clearly demonstrates the promise of the ANN-based approach.

What physical features of the jet are identified by the ANN as the primary characteristics of a top jet? Some insight is provided by the energy deposit patterns of the highest-scoring and lowest-scoring jets, according to the ANN output $\mathcal{O}$, in the top sample. These

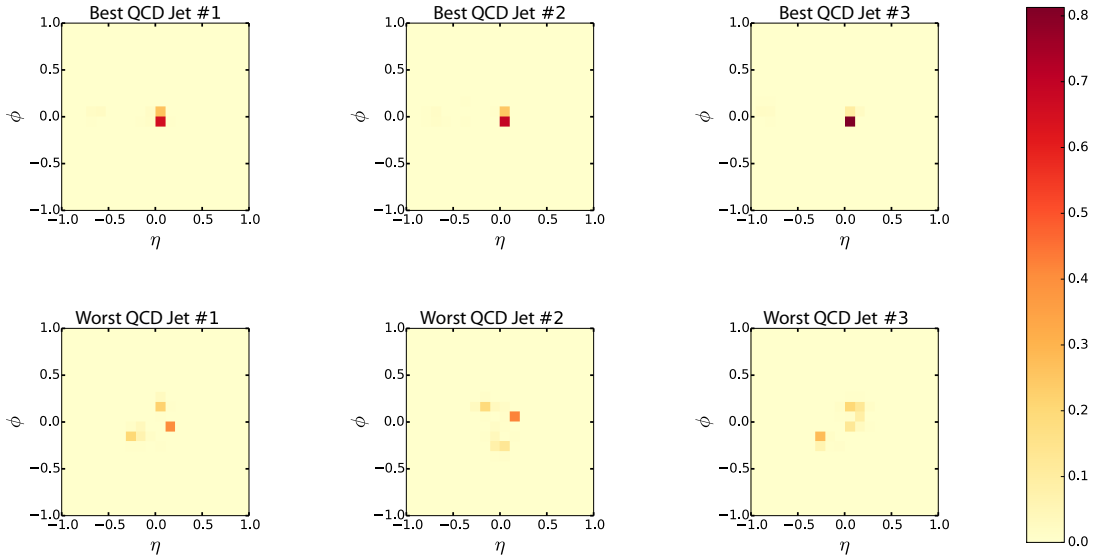

Figure 4.6: Energy deposit patterns for three jets with the lowest (top row) and highest (bottom row) ANN scores in the QCD jet sample with $p_T \in [800, 900]$ GeV.

are shown in Fig. 4.5. It is clear that the jets receiving high scores are characterized by well-defined three-prong structure, with each of the three quarks from top decay forming a well-defined, relatively isolated subjet. The lowest-scoring jets are those where either the quarks are nearly collinear, or one of them is much softer than the other two (in the detector frame). Likewise, the QCD jets receiving the highest scores, and thus most likely to be mis-identified as tops, have well-defined, isolated subjets, while the QCD jets correctly tagged as such do not: see Fig. 4.6.

Tagger	Top		Dijet	
	$p_T \in [500, 600]$	$p_T \in [1100, 1200]$	$p_T \in [500, 600]$	$p_T \in [1100, 1200]$
TOM	0.50	0.52	0.52	0.65
N-sub.	0.59	0.52	0.48	0.31
ATLAS	0.33	0.44	0.42	0.72

Table 4.1: Correlation coefficients between the ANN score and the output of alternative taggers, in a variety of samples.

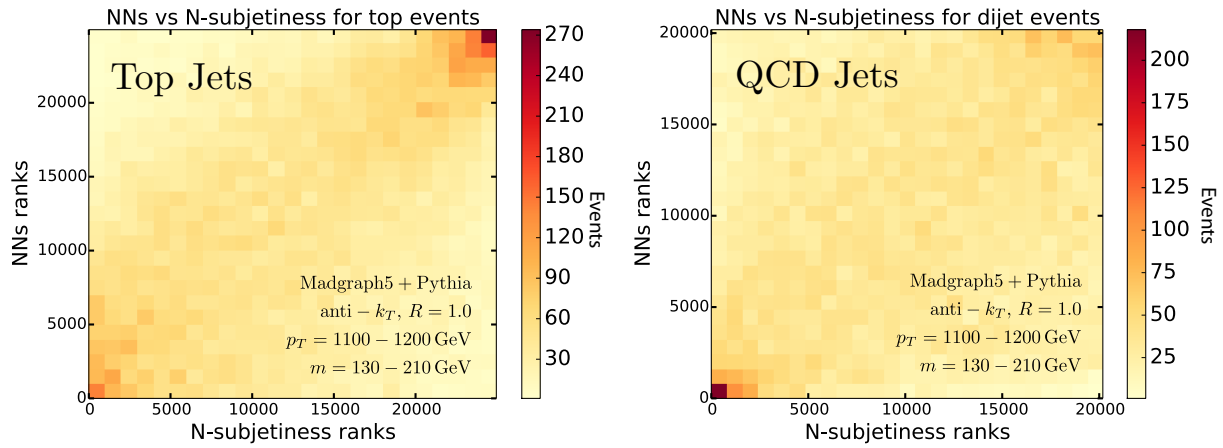

Figure 4.7: Correlation between the rankings of jets according to N-subjettiness (horizontal axis) and ANN score (vertical axis). Left: top sample, $p_T \in [1100, 1200]$ GeV. Right: QCD jet sample, same p_T range. Jets are ranked in order of increasing "topness" for both samples.

To gain further insight, we studied correlations of the ANN scores with other observables used to tag tops. Table 4.1 contains the correlation coefficients between the ANN score and the output of the other taggers in our comparison pool, on a variety of samples used in our analysis. (The correlation coefficients are normalized so that 1.0 indicates perfect correlation and -1.0 perfect anti-correlation, while 0 indicates absence of correlation.) In all cases, we observe significant, though far from perfect, positive correlations, with coefficients ranging from about 0.3 to 0.7. A visual illustration is provided by Fig. 4.7, which shows that the ranking of jets according to the ANN score and the N-subjettiness are indeed correlated, in both top and light-jet samples; correlation plots for all other taggers and p_T ranges look very similar. This should not be surprising since all top taggers to some extent exploit the same physical characteristics of the boosted top jets. Nevertheless, as noted above, ANN systematically outperforms the other taggers in terms of tagging efficiency vs. mistag rates, indicating that the complicated non-linear observable created by the ANN learning process captures the information present in the jet substructure in a more optimal way. In other words, it seems that all taggers find roughly the same sub-

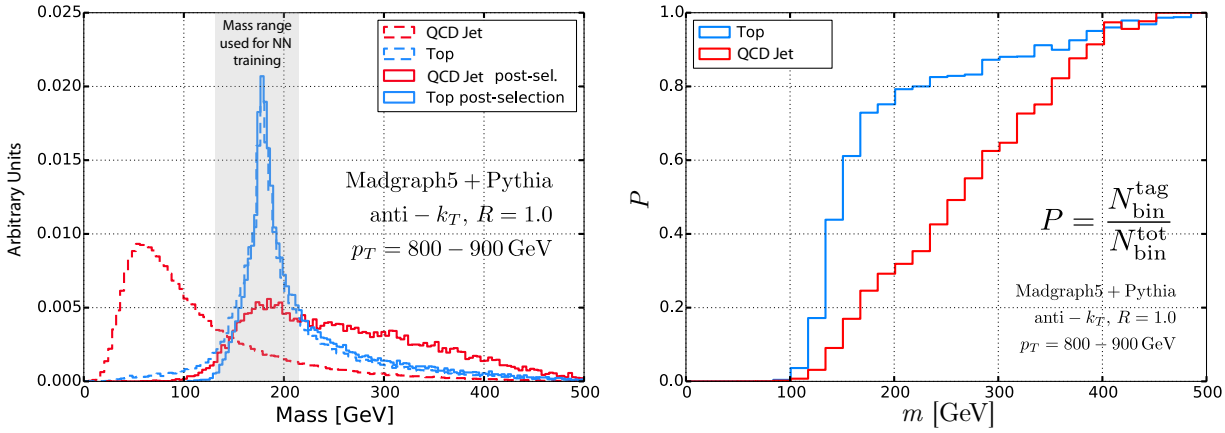

Figure 4.8: Left: Jet mass distributions for top (blue) and dijet (red) samples with $p_T \in [800, 900]$ GeV window, and no mass cut. Dashed lines: all jets; solid lines: jets tagged as tops by the ANN tagger. All distributions are normalized to unit total area. Right: probabilities for a jet in the top (blue) and dijet (red) samples to be tagged as a top jet by the ANN tagger.

set of jets to be "easily classifiable", and all have a very good success rate on this subset. However, the ANN tagger seems to be able to correctly classify a higher fraction of the jets outside of this subset, leading to higher overall success rate.

Another interesting question is how the ANN performance varies with the jet mass. The training samples and test samples in all plots shown so far only contain jets in a $130 \ldots 210$ GeV mass window, where most top jets are expected to lie. We also applied the ANN tagger to the full sample of jets in the $[800, 900]$ GeV p_T range, without the mass cut. The jet mass distributions in this sample, before and after the ANN tagger is applied, as well as the tagging probability as a function of the jet mass, are shown in Fig. 4.8. (The cut on the ANN output used in the figure corresponds to the overall tag efficiency in the $130 \ldots 210$ GeV mass window of 70%.) For jet mass below 130 GeV, the probability of a positive top tag drops rapidly, for both top and QCD jets. This is presumably due to the fact that jets with a clear three-prong structure are unlikely to have a low mass. On the

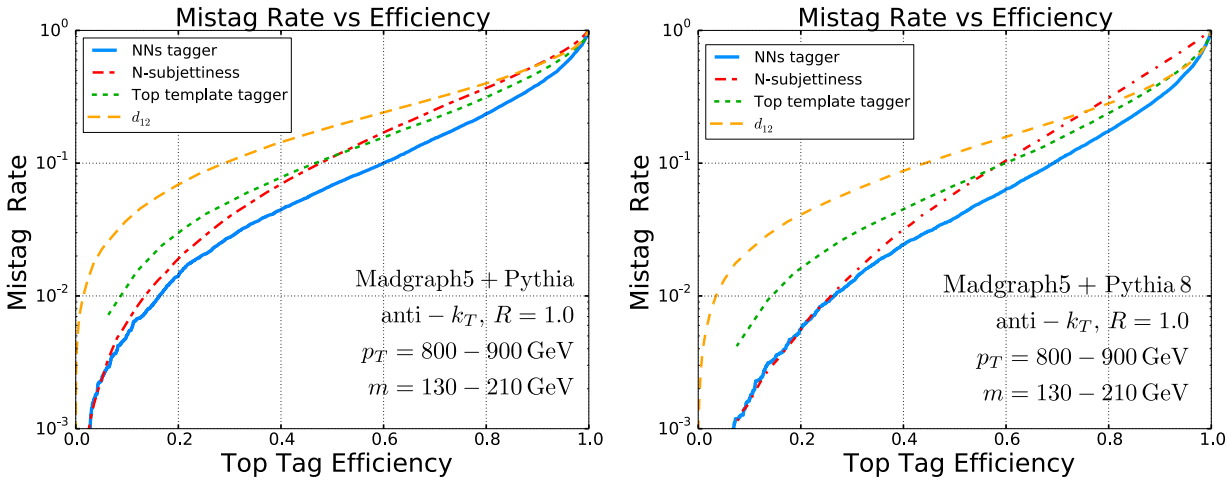

Figure 4.9: Efficiency vs. Mis-tag rate curves for the ANN tagger (blue/solid lines), on jet samples generated with `Pythia 6` (left) and `Pythia 8` (right). For comparison, corresponding curves for three existing top taggers are also shown: d_{12} tagger (yellow/dashed), top template tagger (green/dotted), and N-subjettiness (red/dash-dotted).

other hand, for jet mass above 210 GeV, the probability of a positive top tag is roughly independent of the jet mass. It should also be noted that the tag probability is smooth on the boundaries of the mass window selected for training, indicating that there is no strong dependence on the choice of the training sample. The ability of the ANN tagger to reject jets with low invariant mass may be useful in reducing effects of the pile-up.

The final issue we address is the IR-safety of the ANN output. As any observable in jet physics, the ANN score must be IR-safe (or at least Sudakov-safe [172]) to be useful. Canonically, IR-safety simply requires that the observable be unchanged by exactly collinear $1 \to 2$ patron splitting, or an emission of an infinitely soft gluon. Since neither process affects the energy deposits in calorimeter cells ε_{ab}, and since those energy deposits are the only information used by the ANN, its output is manifestly IR-safe by this definition. As a practical matter, however, one might still worry about the sensitivity of the output to non-perturbative physics involved in splittings at small, but finite, angles, and

emission of gluons with small, but finite, energy. The modeling of this physics in MC generators such as `Pythia` involves approximations with poorly understood systematic errors, and if the ANN output were determined predominantly by features that depend strongly on the showering model, MC studies would clearly be of very limited utility in assessing the ANN performance on real data. To address this concern, we applied the ANN tagger, trained as described above on jet samples showered with `Pythia 6`, to alternative jet samples generated with the same physics inputs but showered with `Pythia 8`. Showering algorithm of `Pythia 8` differs significantly from `Pythia 6`, in that it incorporates a p_T ordered showering as well as an increased number of underlying event modes and the capability to consider two hard processes. We also applied the three taggers in our comparison pool to the same sample. The result is shown in Fig. 4.9. The ANN tagger continues to perform well on test samples generated with a showering model different from the one used in the training set. This indicates that the features ANN uses to classify jets are physical, rather than artifacts of a particular showering model. Moreover, while there is a non-trivial dependence of the efficiency/mis-tag rate curves on the generator, the effect is of the same size for all taggers considered here. In other words, ANN does not appear to be unusually sensitive in this regard.

4.6 Discussion

In this paper, we proposed and explored a new approach to the analysis of jet substructure, specifically top-jet tagging, based on Artificial Neural Network (ANN). The main result of the analysis is captured in Fig. 4.4: the ANN tagger significantly outperforms traditional taggers on the MC "datasets" used in our study. In a sense, this should not come as a surprise: while the ANN uses the same input information as any other tagger, the training procedure constructs a non-linear function of these inputs which is specifi-

cally chosen to maximize its power to classify jets. This maximization takes place on a restricted but extremely broad set of functions, encoded in Fig. 4.1 or Eq. (1.37), and the resulting observable is probably not far away from the theoretical upper limit on classification performance. If this is indeed the case, the ANN can be useful in theoretical studies, serving as a benchmark for other observables used for boosted top tagging.

Being the first study of this novel approach to top tagging, the analysis presented here does not yet fully capture the complexity of the problem in a realistic experimental environment. The very promising results of this analysis strongly motivate further explorations. Some of the important outstanding issues include:

- The jets were extracted from event samples including only leading-order SM processes, $t\bar{t}$ and dijet. Subleading processes need to be included. In spite of their smaller rate, they may have outsize effect on the tagger performance: for example, pure QCD processes with high multiplicity of partons in the final state can create "accidental substructure" [173, 174], and the ANN would need to learn to distinguish it from real top jets.
- Pile-up has not been included in our simulations. While many methods to reduce the effects of pile-up have been suggested [150, 151], their interaction with the ANN tagger needs to be explored.
- Before the method can be applied to real data, concerns about possible MC biases in training the ANN need to be addressed. A preliminary study of this issue suggests that the features that determine the ANN output are not strongly MC-dependent, see Fig. 4.9. However, a more extensive study of this issue is needed, ideally using control/validation samples from real LHC data. In principle, it may even be possible to train the ANN directly on real data, assuming that sufficiently robust training samples can be extracted. This approach would entirely remove concerns about MC biases, and warrants further investigation.

We plan to address some of these issues in future work.

Another important direction is to further improve the tagger performance. A clear limitation of our tagger is that it only uses HCAL information. Other pieces of information are highly relevant for top tagging, the most obvious one being a sub-jet b-tag. This information can certainly be combined with the algorithm presented here to construct an even more powerful tagger. Also, the tagger presented here is based on a rather simple NN architecture and training procedure; more advanced techniques, such as using a convolutional neural network or pre-training the neural network with unsupervised techniques, may result in improved performance.

Finally, while in this paper we focused exclusively on tops, this approach can equally well be applied to other boosted-object jets, such as W and h. It would be interesting to see if performance improvements with respect to traditional taggers can also be achieved in those cases.

In summary, the novel approach to jet tagging based on pattern-recognition techniques, specifically Artificial Neural Networks, shows promise of significant improvements in tagger performance. While the analysis presented in this paper is only the first step, we hope that this approach will eventually become a useful tool in experimental searches for new physics.

CHAPTER 5

CONCLUSION

With the LHC having recently restarted its proton-proton collisions program, we can be hopeful about what lies ahead for particle physics. One of the most attractive possibility is that we will discover some superpartners, thereby solving the hierarchy problem and also potentially the identity of dark matter. However, as long as the LHC finds something new it is safe to say that the field will thrive. This thesis is however my last contribution to physics. Having been part of that community has been the most rewarding and humbling experience of my life. I feel privileged to have had the opportunity to contribute to this field and I have the uttermost respect for my fellow physicists who will keep on pushing the boundary of our knowledge. How can one not be romantic about physics when you contemplate its scope, its history and its aspiration?

APPENDIX A

APPENDIX TO CHAPTER 2: CP-PRESERVING POTENTIAL

In Section 2.3, we presented a list of P- and T-preserving operators in the non-relativistic potential for DM self-interactions. In this appendix we present the additional terms in the effective potential that are generated when parity invariance is relaxed. In addition to $\mathcal{O}_{1,2,3,4,7,8}$, the four operators $\mathcal{O}_{9,10,11,12}$ in (2.13)–(2.15) preserve CP but break parity. For simplicity we consider only the case of self-conjugate DM so that $\mathcal{O}_{10,12}$ are forbidden.

The $\mathcal{O}_9$ term contains no $\mathbf{v}_\perp$ factors and the corresponding potential is

$$V_9 = \frac{\widetilde{g}_9(r)}{4\pi r^3 \Lambda}(\mathbf{s}_1 \times \mathbf{s}_2)\cdot \mathbf{r} \tag{A.1}$$

where $\widetilde{g}_9(r)$ is defined analogously to (2.24).

In order to determine V_{11} we need the Fourier transform of the propagator along the direction tranverse to the exchanged momentum

$$\int \frac{d^3\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q}\cdot\mathbf{r}} \left[\delta^{ij} - \frac{\mathbf{q}_i\mathbf{q}_j}{\mathbf{q}^2}\right] \frac{1}{(\mathbf{q}^2+\mu^2)} = \frac{e^{-\mu r}}{4\pi r}\left[\frac{2}{3}\delta^{ij} + \frac{1}{\mu^2 r^2}\left(3\hat{r}^i\hat{r}^j - \delta^{ij}\right)\left(e^{\mu r} - 1 - \mu r - \frac{\mu^2 r^2}{3}\right)\right] .$$

Contracting this expression with $(\mathbf{s}_1 - \mathbf{s}_2)^i$ and $\mathbf{v}^j$ gives V_{11}. Since the final result is quite involved, we focus on two interesting limits. At distances smaller than the mediator Compton wavelength, $\Lambda^{-1} \ll r \ll \mu^{-1}$, the expression greatly simplifies because

$$\lim_{\mu\to 0}\int \frac{d^3\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q}\cdot\mathbf{r}} \left[\delta^{ij} - \frac{\mathbf{q}_i\mathbf{q}_j}{\mathbf{q}^2}\right] \frac{1}{(\mathbf{q}^2+\mu^2)} = \frac{1}{8\pi r}\left(\delta^{ij} + \hat{r}^i\hat{r}^j\right), \tag{A.2}$$

and hence

$$V_{11} = \frac{1}{8\pi r}\left[(\mathbf{s_1} - \mathbf{s}_2)\cdot\mathbf{v} + (\mathbf{s_1} - \mathbf{s}_2)\cdot\hat{r}(\hat{r}\cdot\mathbf{v})\right] . \tag{A.3}$$

On the other hand, at scales where the mediator mass is important, $r \gg \mu^{-1}$, we have

$$V_{11} = \frac{1}{4\pi r^3 m^2}\left[3(\mathbf{s_1} - \mathbf{s}_2)\cdot\hat{r}\ \ (\hat{r}\cdot\mathbf{v}) - (\mathbf{s_1} - \mathbf{s}_2)\cdot\mathbf{v}\right] . \tag{A.4}$$

where $m^2 = \int d\mu^2 \rho(\mu^2)/\mu^2$.

We stress that the ordering of the various operators in the non-static part of the potential is generically important since $\mathbf{p} = m_\chi \mathbf{v}/2$ is the conjugate coordinate associated with the relative distance, $[\mathbf{r}^i, \mathbf{p}^j] = i\delta^{ij}$.

APPENDIX B

APPENDIX TO CHAPTER 2: SOMMERFELD ENHANCEMENT FOR NON-SINGULAR POTENTIALS

We briefly review the general method to obtain the Sommerfled enhancement [39, 38]. Consider two particles of mass m_χ and center-of-mass momentum $\mathbf{p}$. The ℓ-wave amplitude $A_\ell(\mathbf{p})$ for the annihilation of these two particles under an attractive central potential $V(r)$ can be expressed as a function of a bare amplitude $A_{0,\ell}(\mathbf{q}) = a_{0,\ell} q^\ell$ and a wavefunction $\phi_{\mathbf{p}}(\mathbf{r})$,

$$A_\ell(\mathbf{p}) = \int d\mathbf{r}\, \phi^*_{\mathbf{p}}(\mathbf{r}) \int d\mathbf{q}\, e^{i\mathbf{q}\cdot\mathbf{r}} A_{0,l}(\mathbf{q}). \tag{B.1}$$

The wavefunction $\phi_{\mathbf{p}}(\mathbf{r})$ satisfies the Schrödinger equation,

$$\left(-\frac{1}{2M}\partial^2 + V(r) - \frac{p^2}{2M}\right)\phi_{\mathbf{p}}(\mathbf{r}) = 0, \tag{B.2}$$

where $M = m_\chi/2$ is the reduced mass and $p = Mv$ is the non-relativistic momentum. In general, the potential $V(r)$ can be matrix valued in the space of partial waves, in which case the Schrödinger equation is then a system of coupled differential equations. To solve this equation we decompose the wavefunction $\phi_{\mathbf{p}}(\mathbf{r})$ in partial waves

$$\phi_{\mathbf{p}}(\mathbf{r}) = \frac{(2\pi)^{3/2}}{4\pi p}\sum_\ell (2\ell+1) e^{i\delta_\ell} R_{p,\ell}(r) P_\ell(\hat{\mathbf{p}}\cdot\hat{\mathbf{r}}) \tag{B.3}$$

such that the radial part, $R_{p,\ell}(r)$, satisfies

$$\frac{-1}{2M}\left(\frac{d^2R_{p,\ell}}{dr^2} + \frac{2}{r}\frac{dR_{p,\ell}}{dr} - \frac{\ell(\ell+1)}{r^2}R_{p,\ell}\right) - \left(\frac{p^2}{2M} - V(r)\right)R_{p,\ell} = 0 \tag{B.4}$$

with the completeness relation

$$\int_0^\infty dp R_{p,\ell}(r) R_{p,\ell}(r') = \frac{\delta(r-r')}{r^2}. \tag{B.5}$$

Plugging the partial wave decomposition (B.3) into (B.1) along with $\phi^0_{\mathbf{p}}(\mathbf{r}) = e^{i\mathbf{p}\cdot\mathbf{r}}$ gives

$$A_l(p) = \frac{1}{p}\int_0^\infty r^2 dr R_{p,\ell}(r)\int_0^\infty q dq R^0_{q,\ell}(r) A_{0,\ell}(q). \tag{B.6}$$

From the free solution $R^0_{p,\ell}$ we know that

$$\frac{d^\ell}{dr^\ell} R^0_{q,\ell}(r=0) = \sqrt{\frac{2}{\pi}}\frac{\ell! q^{\ell+1}}{(2\ell+1)!!}. \tag{B.7}$$

Applying the completeness relation (B.6) gives

$$A_\ell(p,p') = \sqrt{\frac{\pi}{2}}\frac{(2\ell+1)!!}{\ell!}\frac{1}{p}\frac{d^\ell}{dr^\ell}R_{p,\ell}(r=0)a_{0,\ell} \tag{B.8}$$

such that the Sommerfeld enhancement for a the ℓ^{th} partial wave is

$$S_l = \left|\sqrt{\frac{\pi}{2}}\frac{(2\ell+1)!!}{\ell!}\frac{1}{p^{\ell+1}}\frac{d^\ell}{dr^\ell}R_{p,\ell}(r=0)\right|^2 \tag{B.9}$$

We thus see that the Sommerfeld enhancement is given by the solution of the Schrödinger equation at the origin.

B.0.1 Numerical algorithm

Refs. [38, 39] provide a method to numerically evaluate the enhancement factor S. The completeness relation (B.5) is valid at long distances,

$$R_{p,\ell}(r)|_{r\to\infty} \to \sqrt{\frac{2}{\pi}}\frac{\sin(pr - \ell\pi/2 + \delta_\ell)}{r}. \tag{B.10}$$

For simplicity, let us work with the dimensionless variable $x = pr$ and the rescaled wave-function $\Phi_{p,\ell}(x) = \frac{xR_{p,\ell}(x)}{Np}$ where N is an arbitrary normalization. Using these variables, the Schrödinger equation takes the form

$$-\Phi_{p,\ell}(x)'' + \left(\mathcal{V}(x) + \frac{\ell(\ell+1)}{x^2} - 1\right)\Phi_{p,\ell}(x) = 0 \tag{B.11}$$

where $\mathcal{V}(x) = \frac{2M}{p^2}V(x/p)$ and we impose the initial conditions

$$\lim_{x\to 0}\Phi_{p,\ell}(x) = x^{\ell+1}. \tag{B.12}$$

From (B.11) and the fact that $\lim_{x\to\infty} \mathcal{V}(x) = 0$, it is clear that in the asymptotically far away region,

$$\Phi_\ell(x)|_{x\to\infty} \to C\sin(x - \ell\pi/2 + \delta_\ell) \tag{B.13}$$

Moreover, to satisfy the asymptotic normalization of $R_{p,\ell}(r)$, we need to fix the normalization $N = \sqrt{\frac{2}{\pi}}\frac{1}{C}$. We can then use $R_{p,\ell} = Np\Phi_l/x$ in (B.9) along with the initial condition to obtain

$$A_\ell(p) = \frac{(2\ell+1)!!}{C} p^\ell a_{0,\ell} = \frac{(2\ell+1)!!}{C} A_{0,\ell}(p) \tag{B.14}$$

so that the Sommerfeld factor is

$$S = \left(\frac{(2\ell+1)!!}{C}\right)^2 \tag{B.15}$$

We thus reduce the calculation of the Sommerfeld enhancement S to the determination of C. This is obtained by numerically solving (B.11) with the initial condition (B.12) and

$$C^2 = \left(\Phi_l(x)^2 + \Phi_l(x - \pi/2)^2\right)|_{x\to\infty}\,. \tag{B.16}$$

B.0.2 Coulomb and Yukawa example

For the Coulomb potential $V(r) = -\alpha/r$, one can obtain an analytic expression for the Sommerfeld enhancement [39, 38],

$$S_\ell = \frac{e^{\pi\alpha/v}\pi\alpha}{v\sinh(\pi\alpha/v)\,\ell!^2} \prod_{s=1}^{\ell}\left(s^2 + \frac{\alpha^2}{v^2}\right) \approx \frac{2\pi}{\ell!^2}\left(\frac{\alpha}{v}\right)^{2\ell+1} \tag{B.17}$$

where the approximation holds for large α/v. There exists no simple analytical expression for the enhancement from a Yukawa potential $V(r) = -\alpha e^{-\mu r}/r$, but one can easily evaluate it numerically using the method presented, see Fig. (B.1). The presence of resonances can be explained by bound states [42, 175].

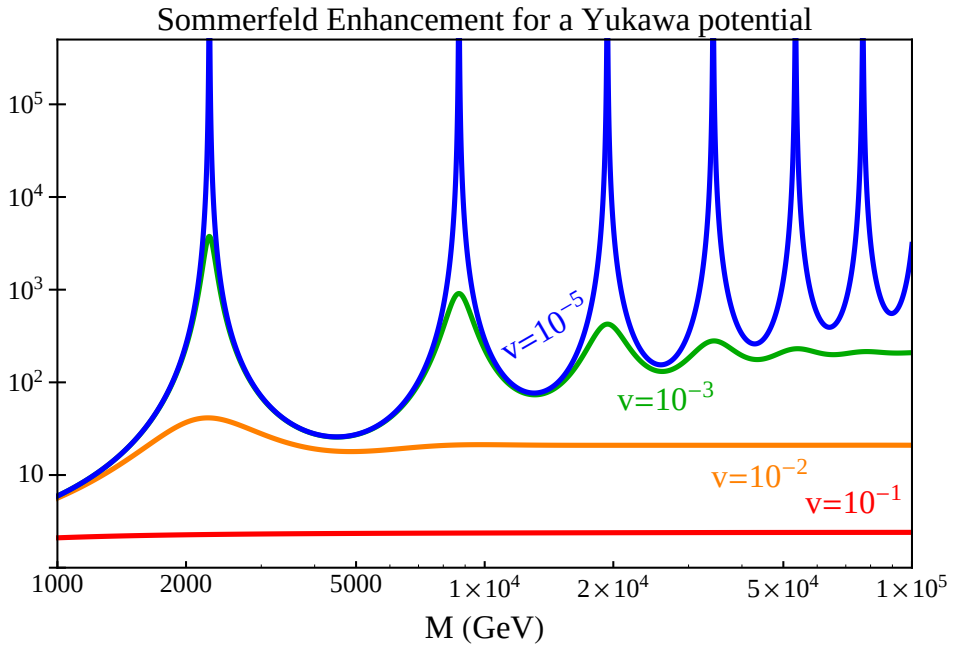

Figure B.1: Numerical evaluation of the Sommerfeld enhancement factor as a function of the dark matter reduced mass M for a range of relative velocities. The mediator mass is fixed to 90 GeV and $\alpha = 1/30$.

APPENDIX C

APPENDIX TO CHAPTER 2: BOX APPROXIMATION

We have shown that bound state resonances can generate large Sommerfeld enhancements. In this appendix we adapt the procedure used in [42] to quantitatively understand these resonances. In [42], it was shown that the a reasonable approximation for the Yukawa potential is a flat potential well whose width is determined by the characteristic length scale of the interaction, $r_0 = 1/m_\varphi$,

$$V_{\text{box}}(r) = -U_0 \Theta(r_0 - r). \tag{C.1}$$

The depth of the rectangular well U_0 is fixed by requiring that the box approximation matches the Yukawa potential at $r = r_0$,

$$V_{\text{box}}(r) = -\frac{\alpha m}{e} \Theta\left(\frac{1}{m} - r\right). \tag{C.2}$$

This approximate is constructed to capture only the qualitative behavior of the full potential and is not a detailed matching to an effective theory. Observe that this analysis agrees with the fact that the Coulomb limit ($m_\varphi \to 0$) does not have resonances: this potential has no natural length scale for constructing the rectangular well.

C.0.3 Application to $V \sim r^{-3}$

We adapt this procedure to the singular $1/r^3$ potential,

$$V(r) = \frac{-\alpha}{f^2} \frac{1}{r^3}. \tag{C.3}$$

The natural length scale of the problem is the dimensionful scale of the coupling, $r_0 = \sqrt{\alpha}/f$. In principle there is also a scale set from the exponential term $e^{-m_\varphi r}$, but for UV models with $m_\varphi \ll f$ this contribution is negligible. This reflects the fact that the resonant behavior of singular potentials in this limit do not depend strongly on the specific value of the mediator mass m_φ.

This simple box potential approximation provides an estimate for the upper bound of Sommerfeld enhancement coming from resonances in a singular potential. The solution to the $\ell = 0$ Schrödinger equation inside the box ($r < r_0$) is

$$\phi(pr < pr_0)|_p = \frac{\sin(\kappa p r)}{\kappa}, \tag{C.4}$$

where $\kappa p = \sqrt{p^2 + 2U_0 M}$. Outside the box, $r > r_0$, there is effective no potential so that

$$\phi(pr > pr_0)_p = C \sin(pr + \delta). \tag{C.5}$$

C is determined by requiring continuity at r_0 so that the enhancement is

$$\begin{aligned} S &= \left[\cos^2(\kappa p r_0) + \frac{\sin^2(\kappa p r_0)}{\kappa^2}\right]^{-1} \\ &\approx \left[\cos^2\left(r_0\sqrt{2U_0 M}\right) + \frac{p^2}{2MU_0}\sin^2\left(r_0\sqrt{2U_0 M}\right)\right]^{-1}, \end{aligned} \tag{C.6}$$

where we use the non-relativistic approximation $p^2 \ll U_0 M$. Observe that the prefactor of the sine term is small so that S becomes large when the cosine vanishes. In other words, this expression maximized when $r_0\sqrt{2U_0 M} = (2n+1)\pi/2$ with

$$S_{\text{max}} \approx \frac{2MU_0}{p^2} = \frac{(2n+1)^2\pi^2}{4r_0^2 p^2}. \tag{C.7}$$

This peak is exactly the resonance when the pair of dark matter particles forms a bound state. Note that this approximation is independent of the depth of the rectangular well, U_0.

It is straightforward to generalize these expressions for an arbitrary orbital angular momenta, ℓ, by including the angular barrier to the box potential and applying the appropriate boundary conditions. One obtains

$$S_\ell = \left(\frac{\pi\left[(2\ell')!!\right]^2 \widetilde{\kappa}^{2\ell'}}{2^{2\ell'+1}\Gamma(\ell'+1)^2}\right) \frac{\left[Y_{\ell'}(pr_0) - \cot(\delta) J_{\ell'}(pr_0)\right]^2}{\left[1 + \cot^2(\delta)\right] J_{\ell'}^2(\widetilde{\kappa} p r_0)} \tag{C.8}$$

where $\ell' = \ell + \frac{1}{2}$ and $\widetilde{\kappa}^2 = 2MU_0/p^2$. The qualitative scaling behavior of the resonance can be seen by setting $\cot(\delta) = 0$, and assuming that $pr_0 \ll 1$ so that

$$S_{\text{max}} \sim \frac{1}{\left(p^2 r_0^2\right)^{2\ell+1}} \sim \frac{1}{v^{4\ell+2}}. \tag{C.9}$$

C.0.4 Dimensional analysis

To estimate the Sommerfeld enhancement off resonance one must estimate U_0. We use the assumption that the UV physics encoded in U_0 does not significantly change the IR potential so that the height of the square well U_0 is well approximated by the value of the singular potential at the cutoff scale,

$$U_0 \sim \frac{f}{\alpha^{1/2}} \sim \frac{1}{r_0}, \tag{C.10}$$

so that for $\ell = 0$, the Sommerfeld enhancement is approximately

$$S \approx \left[\cos^2\left(\sqrt{\frac{2M\alpha^{1/2}}{f}} \right) + \frac{p^2\alpha^{1/2}}{2Mf} \sin^2\left(\sqrt{\frac{2M\alpha^{1/2}}{f}} \right) \right]^{-1}. \tag{C.11}$$

An estimate for the parameters required to hit a resonance without tuning is thus

$$M_{\text{res}} \sim \frac{1}{r_0} \sim \frac{f}{\alpha^{1/2}}, \tag{C.12}$$

which, for most cases, lies at the boundary of the range of the theory's validity.

APPENDIX D

APPENDIX TO CHAPTER 3: COLLIDER BOUNDS

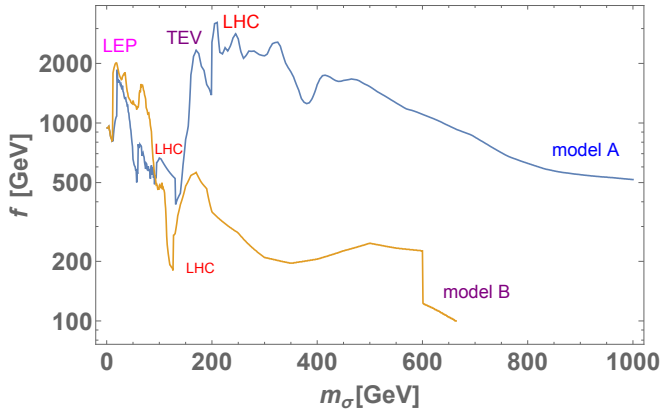

Figure D.1: 95% C.L. collider exclusion limit on the scale of conformal symmetry breaking, f, with respect to m_σ for our benchmark models A and B.

As mentioned in the main text, in addition to the direct detection bounds there are also collider bounds from the LHC and earlier experiments. The dilaton (roughly) mimics a Higgs boson, with couplings to massive SM fields suppressed by the factor v/f compared to that of the Higgs and couplings to massless gauge bosons that involve contributions from the matter content of the conformal sector. Collider bounds on the dilaton can thus be obtained by recasting the results of direct production limits from Higgs boson searches. We use the HiggsBound [176, 177, 178] code version 4.1.2, that incorporates all the currently available experimental analyses from LEP, the Tevatron, and the LHC [176, 177, 178].

The resulting collider bounds on the conformal symmetry breaking scale f as a function of the dilaton mass is presented in Fig. D.1 for the two benchmark models A and B defined in Sec. 3.3. In obtaining these bounds we assumed, for simplicity, no invisible decay channels for the dilaton. We can see that the collider bounds are strongly model dependent: model A has a large coupling to gluons, and thus is very strongly constrained

throughout the parameter space relevant for LHC kinematics. Model B has small couplings to gluons and photons, and is only weakly constrained for dilaton masses above 200 GeV.

The resulting bound on f can be turned into a bound on m_χ using Fig. 3.2. For example the $f \gtrsim 2$ TeV bound for $m_\sigma \lesssim 400$ GeV in model A implies $m_\chi \gtrsim 300$ GeV, with the exception for a narrow resonance region.

APPENDIX E

APPENDIX TO CHAPTER 3: ADDITIONAL ANNIHILATION CHANNELS

E.0.5 Scalar dark matter

Annihilation to fermions of mass m_ψ:

$$\sigma v(\chi\chi \to \bar{\psi}\psi) \simeq \frac{m_\psi^2 m_\chi \left(m_\chi^2 - m_\psi^2\right)^{3/2}}{\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.1}$$

Annihilation to a real scalar of mass m_h:

$$\sigma v(\chi\chi \to hh) \simeq \frac{m_\chi m_h^4 \sqrt{m_\chi^2 - m_h^2}}{4\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.2}$$

While annihilation to photons is negligible for the relic abundance calculation, it is important for indirect detection. We get

$$\sigma v(\chi\chi \to \gamma\gamma) \simeq \frac{3m_\chi^6 \alpha_{\text{EM}}^2 c_{\text{EM}}^2}{16\pi^3 f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.3}$$

where c_{EM} encodes the coupling of photons to dilaton:

$$c_{\text{EM}} = \left(F_W(x_W) - \sum_f N_c Q_f^2 F_f(x_f) + b_{\text{IR}}^{(\text{EM})} - b_{\text{UV}}^{(\text{EM})}\right) \tag{E.4}$$

$$x_i = \frac{m_i^2}{m_\chi^2} \tag{E.5}$$

$$F_W(x) = 2 + 3x + 3x(2-x)f(x) \tag{E.6}$$

$$F_f(x) = 2x[1 + (1-x)f(x)] \tag{E.7}$$

$$f(x) = \begin{cases} \arcsin(1/\sqrt{x})^2 & : x \geq 1 \\ -\frac{1}{4}\left[\log\left(\frac{1+\sqrt{x-1}}{1-\sqrt{x-1}}\right) - i\pi\right]^2 & : x < 1 \end{cases} \tag{E.8}$$

E.0.6 Fermion dark matter

Annihilation to fermions of mass m_ψ:

$$\sigma v(\bar{\chi}\chi \to \bar{\psi}\psi) \simeq v^2 \frac{m_\psi^2 m_\chi \left(m_\chi^2 - m_\psi^2\right)^{3/2}}{8\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.9}$$

Annihilation to a real scalar of mass m_h:

$$\sigma v(\bar{\chi}\chi \to hh) \simeq v^2 \frac{m_\chi m_h^4 \sqrt{m_\chi^2 - m_h^2}}{32\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.10}$$

E.0.7 Vector dark matter

Annihilation to fermions of mass m_ψ:

$$\sigma v(\chi\chi \to \bar{\psi}\psi) \simeq \frac{m_\psi^2 m_\chi \left(m_\chi^2 - m_\psi^2\right)^{3/2}}{3\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.11}$$

Annihilation to a real scalar of mass m_h:

$$\sigma v(\chi\chi \to hh) \simeq \frac{m_\chi m_h^4 \sqrt{m_\chi^2 - m_h^2}}{12\pi f^4 |4m_\chi^2 - m_\sigma^2 - i\text{Im}\left(\Pi(4m_\chi^2)\right)|^2} \tag{E.12}$$

E.1 Dilaton decay channels

For decay to a real scalar of mass m_h we get

$$\Gamma_\sigma(\sigma \to hh) = \frac{m_h^4}{8\pi m_\sigma f^2} \sqrt{1 - \frac{4m_h^2}{m_\sigma^2}} \tag{E.13}$$

For decay to fermions we get

$$\Gamma_\sigma(\sigma \to \bar{\psi}\psi) = \frac{m_\sigma m_\psi^2}{8\pi f^2} \left(1 - \frac{4m_\psi^2}{m_\sigma^2}\right)^{3/2} \tag{E.14}$$

APPENDIX F

APPENDIX TO CHAPTER 4: A BRIEF DESCRIPTION OF EXISTING TOP TAGGERS

For the purpose of comparison of the ANN tagger to the existing algorithms, we have chosen three existing methods, each one exploiting a different approach to boosted top tagging. In the following list, we give a brief description of the algorithms and the parameters we use for the analysis, while we refer the reader to the references within for detailed discussions.

- **Template Overlap Method (TOM)**: TOM [153, 154, 155, 156] is a jet substructure algorithm which aims to match the energy distribution of a fat jet to a partonic structure which models the decay of a heavy boosted particle. TOM algorithm proceeds by comparing libraries of kinematically allowed parton level decays of massive particles ("templates") to the energy distribution of a fat jet. The quality of a match is quantified by the overlap function Ov, which minimises the difference between the parton transverse momenta and the amount of p_T deposited in small angular regions around the template patrons ("template sub cones"). An $Ov \sim 1$ score signals a top like jet, while a $Ov \sim 0$ is characteristic of light QCD jets. Here we use the `TemplateTagger v.1.0` [179] implementation of the TOM algorithm.

  There are many ways generation of template libraries can be implemented. For simplicity and processing speed, here we consider templates at fixed total transverse momentum matched to the mid-point in each fat jet p_T bin of the event samples ($e.g.$ 550 GeV for fat jet $p_T = 500 - 600$ GeV). We generate the template states using a sequential scan of 40 steps in η, ϕ over the angular region of $R = 1.0$ around the fat jet axis. We match the template libraries to the energy distribution of the fat jet using fixed template sub cones of size $r_3 = 0.1, 0.15, 0.2$ for template $p_T = 1150, 850, 550$ GeV respectively, while we allow for the template resolution

parameter $\sigma_a = p_{T,a}/3$, where $p_{T,a}$ is the transverse momentum of an individual template parton.

- **N-subjettiness:** Perhaps the most notable example of a "prong" tagger is N-subjettiness [152, 27]. The algorithm is based on calculating moments τ_N, which serve as estimates of how well the jet energy distribution can be divided into N regions. The τ_N are calculated by minimizing the p_T weighted distances between calorimeter energy depositions and trial axes which divide the distribution into N regions, over the space of possible axis configurations. The N-subjettiness tagger used in our comparisons is the version publicly available on `HepForge`.[1]

  For the purpose of top tagging the most useful observable is typically the ratio τ_3/τ_2, where a high score means that a jet distribution is described better by a three prong configuration. Conversely, a low τ_3/τ_2 score is characteristic of two prong jets. Note that in the analysis of this paper we used the angular weight exponent $\beta = 1$ in calculations of τ_N moments, as suggested in Ref. [27].

- **ATLAS top tagger**: Jet clustering history can provide useful insight into jet substructure. A notable example is the ATLAS top tagger [180] which utilises the differences between the top and light jets in the last step of jet clustering. The observable ATLAS uses is d_{12}, the value of the k_T norm at the clustering step which goes from two subjects to one final jet. The d_{12} observable is sensitive to the dynamics of hard splittings within the fat jet. The highly asymmetric splittings of typical light jets tend to be characterised by low values of d_{12} with a distribution which falls off sharply with the increase in d_{12}, while we expect typical top jets to be characterised by $d_{12} \sim m_t^2/4$.

  In addition to d_{12}, ATLAS also imposes a lower cut on the trimmed jet mass of $m_j > 130$ GeV. Unless otherwise noted, here we omit the lower mass cut as the data samples we use for comparison are already restricted to a jet mass window in Eq. (4.1).

[1] See `http://fastjet.hepforge.org/contrib/contents/latest.html`

BIBLIOGRAPHY

[1] **LUX** Collaboration, D. Akerib *et. al.*, *The Large Underground Xenon (LUX) Experiment*, *Nucl.Instrum.Meth.* **A704** (2013) 111–126, [arXiv:1211.3788].

[2] K. Griest and M. Kamionkowski, *Unitarity limits on the mass and radius of dark-matter particles*, *Phys. Rev. Lett.* **64** (1990), no. 6 615.

[3] J. Einasto, *Dark Matter*, *Braz.J.Phys.* **43** (2013) 369–374, [arXiv:1308.2534].

[4] V. C. Rubin, *One hundred years of rotating galaxies*, *Publications of the Astronomical Society of the Pacific* **112** (2000), no. 772 747–750.

[5] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, *A direct empirical proof of the existence of dark matter*, astro-ph/0608407v1.

[6] W. Hu and S. Dodelson, *Cosmic microwave background anisotropies*, astro-ph/0110414v1.

[7] **Planck** Collaboration, P. Ade *et. al.*, *Planck 2013 results. I. Overview of products and scientific results*, *Astron.Astrophys.* **571** (2014) A1, [arXiv:1303.5062].

[8] S. Dodelson, *Modern Cosmology*. Academic Press, 1 ed., 2003.

[9] E. Kolb and M. Turner, *The Early Universe*. Westview Press, 1994.

[10] **Planck Collaboration** Collaboration, P. Ade *et. al.*, *Planck 2013 results. XVI. Cosmological parameters*, *Astron.Astrophys.* (2014) [arXiv:1303.5076].

[11] J. Terning, *Modern Supersymmetry: Dynamics and Duality*. Oxford University Press, USA, 2006.

[12] M. Frigerio, A. Pomarol, F. Riva, and A. Urbano, *Composite Scalar Dark Matter*, *JHEP* **1207** (2012) 015, [arXiv:1204.2808].

[13] R. Schnee, *Introduction to dark matter experiments*, arXiv:1101.5205.

[14] J. Carr, G. Lamanna, and J. Lavalle, *Indirect detection of dark matter*, *Rept. Prog. Phys.* **69** (2006) 2475–2512.

[15] C. Weniger, *A Tentative Gamma-Ray Line from Dark Matter Annihilation at the Fermi Large Area Telescope*, arXiv:1204.2797.

[16] **PAMELA** Collaboration, O. Adriani *et. al.*, *An anomalous positron abundance in cosmic rays with energies 1.5.100 GeV*, *Nature* **458** (2009) 607–609, [arXiv:0810.4995].

[17] S. Tulin, H.-B. Yu, and K. M. Zurek, *Beyond Collisionless Dark Matter: Particle Physics Dynamics for Dark Matter Halo Structure*, *Phys.Rev.* **D87** (2013), no. 11 115007, [arXiv:1302.3898].

[18] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, *A Theory of Dark Matter*, *Phys.Rev.* **D79** (2009) 015014, [arXiv:0810.0713].

[19] B. Bellazzini, C. Csaki, J. Hubisz, J. Serra, and J. Terning, *A Higgslike Dilaton*, *Eur.Phys.J.* **C73** (2013) 2333, [arXiv:1209.3299].

[20] W. D. Goldberger, B. Grinstein, and W. Skiba, *Distinguishing the Higgs boson from the dilaton at the Large Hadron Collider*, *Phys.Rev.Lett.* **100** (2008) 111802, [arXiv:0708.1463].

[21] S. Marsland, *Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC, 1st ed., 2009.

[22] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1st ed., 1996.

[23] F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain.*, *Psychological review* **65** (1958), no. 6 386.

[24] P. Werbos, *Beyond regression: new tools for prediction and analysis in the behavioral sciences.*, *Thesis (Ph. D.), Harvard University* (1975).

[25] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, *arXiv preprint arXiv:1207.0580* (2012).

[26] G. Hinton, S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets*, *Neural computation* **18** (2006), no. 7 1527–1554.

[27] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **1202** (2012) 093, [arXiv:1108.2701].

[28] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **1502** (2015) 118, [arXiv:1407.5675].

[29] "Kaggle: Higgs boson machine learning challenge." https://www.kaggle.com/c/higgs-boson. Accessed: 2015-03-25.

[30] B. Bellazzini, M. Cliche, and P. Tanedo, *Effective theory of self-interacting dark matter*, *Phys.Rev.* **D88** (2013), no. 8 083506, [arXiv:1307.1129].

[31] **Fermi LAT Collaboration** Collaboration, M. Ackermann *et. al.*, *Measurement of separate cosmic-ray electron and positron spectra with the Fermi Large Area Telescope*, *Phys.Rev.Lett.* **108** (2012) 011103, [arXiv:1109.0521].

[32] **AMS Collaboration** Collaboration, M. Aguilar *et. al.*, *First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5350 GeV*, *Phys.Rev.Lett.* **110** (2013), no. 14 141102.

[33] T. Bringmann, X. Huang, A. Ibarra, S. Vogl, and C. Weniger, *Fermi LAT Search for Internal Bremsstrahlung Signatures from Dark Matter Annihilation*, *JCAP* **1207** (2012) 054, [arXiv:1203.1312].

[34] M. Su and D. P. Finkbeiner, *Strong Evidence for Gamma-ray Line Emission from the Inner Galaxy*, arXiv:1206.1616.

[35] A. Rajaraman, T. M. Tait, and D. Whiteson, *Two Lines or Not Two Lines? That is the Question of Gamma Ray Spectra*, *JCAP* **1209** (2012) 003, [arXiv:1205.4723].

[36] **On Behalf of the Fermi-LAT Collaboration** Collaboration, E. Bloom *et. al.*, *Search of the Earth Limb Fermi Data and Non-Galactic Center Region Fermi Data for Signs of Narrow Lines*, arXiv:1303.2733.

[37] **LAT Collaboration** Collaboration, M. Ackermann *et. al.*, *Fermi LAT Search for Dark Matter in Gamma-ray Lines and the Inclusive Photon Spectrum*, *Phys.Rev.* **D86** (2012) 022002, [arXiv:1205.2739].

[38] R. Iengo, *Sommerfeld enhancement: General results from field theory diagrams*, *JHEP* **0905** (2009) 024, [arXiv:0902.0688].

[39] R. Iengo, *Sommerfeld enhancement for a Yukawa potential*, arXiv:0903.0317.

[40] S. Cassel, *Sommerfeld factor for arbitrary partial wave processes*, *J.Phys.* **G37** (2010) 105009, [arXiv:0903.5307].

[41] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, *Non-perturbative effect on dark*

matter annihilation and gamma ray signature from galactic center, *Phys.Rev.* **D71** (2005) 063528, [hep-ph/0412403].

[42] M. Lattanzi and J. I. Silk, *Can the WIMP annihilation boost factor be boosted by the Sommerfeld enhancement?*, *Phys.Rev.* **D79** (2009) 083523, [arXiv:0812.0360].

[43] E. Braaten and H. W. Hammer, *Universal Two-body Physics in Dark Matter near an S-wave Resonance*, arXiv:1303.4682.

[44] D. N. Spergel and P. J. Steinhardt, *Observational evidence for self-interacting cold dark matter*, *Phys. Rev. Lett.* **84** (2000) 3760–3763, [astro-ph/9909386].

[45] R. Dave, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt, *Halo properties in cosmological simulations of selfinteracting cold dark matter*, *Astrophys.J.* **547** (2001) 574–589, [astro-ph/0006218].

[46] M. Rocha, A. H. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, *et. al.*, *Cosmological Simulations with Self-Interacting Dark Matter I: Constant Density Cores and Substructure*, *Mon.Not.Roy.Astron.Soc.* **430** (2013) 81–104, [arXiv:1208.3025].

[47] A. H. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, *Cosmological Simulations with Self-Interacting Dark Matter II: Halo Shapes vs. Observations*, arXiv:1208.3026.

[48] J. F. Navarro, V. R. Eke, and C. S. Frenk, *The cores of dwarf galaxy halos*, astro-ph/9610187.

[49] X. Chu, T. Hambye, and M. H. Tytgat, *The Four Basic Ways of Creating Dark Matter Through a Portal*, *JCAP* **1205** (2012) 034, [arXiv:1112.0493].

[50] M. R. Buckley and P. J. Fox, *Dark Matter Self-Interactions and Light Force Carriers*, *Phys.Rev.* **D81** (2010) 083522, [arXiv:0911.3898].

[51] J. L. Feng, M. Kaplinghat, and H.-B. Yu, *Halo Shape and Relic Density Exclusions of Sommerfeld-Enhanced Dark Matter Explanations of Cosmic Ray Excesses*, *Phys.Rev.Lett.* **104** (2010) 151301, [arXiv:0911.0422].

[52] H. Georgi, *Unparticle physics*, *Phys.Rev.Lett.* **98** (2007) 221601, [hep-ph/0703260].

[53] H. Georgi and Y. Kats, *Unparticle self-interactions*, *JHEP* **1002** (2010) 065, [arXiv:0904.1962].

[54] B. Bellazzini, C. Csaki, J. Hubisz, J. Shao, and P. Tanedo, *Goldstone Fermion Dark Matter*, arXiv:1106.2162.

[55] S. Chang, N. Weiner, and I. Yavin, *Magnetic Inelastic Dark Matter*, *Phys.Rev.* **D82** (2010) 125011, [arXiv:1007.4200].

[56] F. Ferrer and M. Nowakowski, *Higgs and Goldstone bosons mediated long range forces*, *Phys.Rev.* **D59** (1999) 075009, [hep-ph/9810550].

[57] S. D. Hsu and P. Sikivie, *Long range forces from two neutrino exchange revisited*, *Phys.Rev.* **D49** (1994) 4951–4953, [hep-ph/9211301].

[58] G. Feinberg, J. Sucher, and C. Au, *THE DISPERSION THEORY OF DISPERSION FORCES*, *Phys.Rept.* **180** (1989) 83.

[59] B. A. Dobrescu and I. Mocioiu, *Spin-dependent macroscopic forces from new particle exchange*, *JHEP* **0611** (2006) 005, [hep-ph/0605342].

[60] P. F. Bedaque, M. I. Buchoff, and R. K. Mishra, *Sommerfeld enhancement from Goldstone pseudo-scalar exchange*, *JHEP* **0911** (2009) 046, [arXiv:0907.0235].

[61] J. Fan, M. Reece, and L.-T. Wang, *Non-relativistic effective theory of dark matter direct detection*, *JCAP* **1011** (2010) 042, [arXiv:1008.1591].

[62] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, and Y. Xu, *The Effective Field Theory of Dark Matter Direct Detection*, *JCAP* **1302** (2013) 004, [arXiv:1203.3542].

[63] W. Frank, D. Land, and R. Spector, *Singular potentials*, *Rev.Mod.Phys.* **43** (1971) 36–98.

[64] G. Lepage, *How to renormalize the Schrodinger equation*, nucl-th/9706029.

[65] S. Beane, P. F. Bedaque, L. Childress, A. Kryjevski, J. McGuire, *et. al.*, *Singular potentials and limit cycles*, *Phys.Rev.* **A64** (2001) 042103, [quant-ph/0010073].

[66] S.-H. Oh, W. de Blok, E. Brinks, F. Walter, and J. Kennicutt, Robert C., *Dark and luminous matter in THINGS dwarf galaxies*, arXiv:1011.0899.

[67] R. K. de Naray and K. Spekkens, *Do Baryons Alter the Halos of Low Surface Brightness Galaxies?*, *Astrophys.J.* **741** (2011) L29, [arXiv:1109.1288].

[68] R. A. Flores and J. R. Primack, *Observational and theoretical constraints on singular dark matter halos*, *Astrophys.J.* **427** (1994) L1–4, [astro-ph/9402004].

[69] J. D. Simon, A. D. Bolatto, A. Leroy, L. Blitz, and E. L. Gates, *High-resolution measurements of the halos of four dark matter-dominated galaxies: Deviations from a universal density profile*, *Astrophys.J.* **621** (2005) 757–776, [astro-ph/0412035].

[70] G. R. Blumenthal, S. Faber, R. Flores, and J. R. Primack, *Contraction of dark matter galactic halos due to baryonic infall*, *The Astrophysical Journal* **301** (1986) 27–34.

[71] O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai, *Response of dark matter halos to condensation of baryons: Cosmological simulations and improved adiabatic contraction model*, *Astrophys.J.* **616** (2004) 16–26, [astro-ph/0406247].

[72] P. B. Tissera, S. D. White, S. Pedrosa, and C. Scannapieco, *Dark matter response to galaxy formation*, arXiv:0911.2316.

[73] M. Vogelsberger, J. Zavala, and A. Loeb, *Subhaloes in Self-Interacting Galactic Dark Matter Haloes*, *Mon.Not.Roy.Astron.Soc.* **423** (2012) 3740, [arXiv:1201.5892].

[74] T. Sawala, Q. Guo, C. Scannapieco, A. Jenkins, and S. D. White, *What is the (Dark) Matter with Dwarf Galaxies?*, arXiv:1003.0671.

[75] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, *Too big to fail? The puzzling darkness of massive Milky Way subhaloes*, *Mon.Not.Roy.Astron.Soc.* **415** (2011) L40, [arXiv:1103.0007].

[76] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, *The Milky Way's bright satellites as an apparent failure of LCDM*, *Mon.Not.Roy.Astron.Soc.* **422** (2012) 1203–1218, [arXiv:1111.2048].

[77] B. Moore, S. Ghigna, F. Governato, G. Lake, T. R. Quinn, *et. al.*, *Dark matter substructure within galactic halos*, *Astrophys.J.* **524** (1999) L19–L22, [astro-ph/9907411].

[78] A. A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, *Where are the missing Galactic satellites?*, *Astrophys.J.* **522** (1999) 82–92, [astro-ph/9901240].

[79] I. M. Shoemaker, *Constraints on Dark Matter Protohalos in Effective Theories and Neutrinophilic Dark Matter*, arXiv:1305.1936.

[80] J. Miralda-Escude, *A test of the collisional dark matter hypothesis from cluster lensing*, astro-ph/0002050.

[81] M. Oguri, M. Takada, N. Okabe, and G. P. Smith, *Direct measurement of dark matter*

halo ellipticity from two-dimensional lensing shear maps of 25 massive clusters, *Mon.Not.Roy.Astron.Soc.* **405** (2010) 2215–2230, [arXiv:1004.4214].

[82] D. P. Finkbeiner, L. Goodenough, T. R. Slatyer, M. Vogelsberger, and N. Weiner, *Consistent Scenarios for Cosmic-Ray Excesses from Sommerfeld-Enhanced Dark Matter Annihilation*, *JCAP* **1105** (2011) 002, [arXiv:1011.3082].

[83] S. Hannestad and T. Tram, *Sommerfeld Enhancement of DM Annihilation: Resonance Structure, Freeze-Out and CMB Spectral Bound*, *JCAP* **1101** (2011) 016, [arXiv:1008.1511].

[84] J. Hisano, M. Kawasaki, K. Kohri, T. Moroi, K. Nakayama, *et. al.*, *Cosmological constraints on dark matter models with velocity-dependent annihilation cross section*, *Phys.Rev.* **D83** (2011) 123511, [arXiv:1102.4658].

[85] A. Loeb and N. Weiner, *Cores in Dwarf Galaxies from Dark Matter with a Yukawa Potential*, *Phys.Rev.Lett.* **106** (2011) 171302, [arXiv:1011.6374].

[86] L. Bergstrom, J. Edsjo, and G. Zaharijas, *Dark matter interpretation of recent electron and positron data*, *Phys.Rev.Lett.* **103** (2009) 031103, [arXiv:0905.0333].

[87] J. L. Feng, M. Kaplinghat, and H.-B. Yu, *Sommerfeld Enhancements for Thermal Relic Dark Matter*, *Phys.Rev.* **D82** (2010) 083525, [arXiv:1005.4678].

[88] T. Cohen, M. Lisanti, T. R. Slatyer, and J. G. Wacker, *Illuminating the 130 GeV Gamma Line with Continuum Photons*, *JHEP* **1210** (2012) 134, [arXiv:1207.0800].

[89] J. Fan and M. Reece, *A Simple Recipe for the 111 and 128 GeV Lines*, arXiv:1209.1097.

[90] C. Jackson, G. Servant, G. Shaughnessy, T. M. Tait, and M. Taoso, *Higgs in Space!*, *JCAP* **1004** (2010) 004, [arXiv:0912.0004].

[91] C. Jackson, G. Servant, G. Shaughnessy, T. M. P. Tait, and M. Taoso, *Gamma Rays from Top-Mediated Dark Matter Annihilations*, arXiv:1303.4717.

[92] C. Jackson, G. Servant, G. Shaughnessy, T. M. P. Tait, and M. Taoso, *Gamma-ray lines and One-Loop Continuum from s-channel Dark Matter Annihilations*, arXiv:1302.1802.

[93] J. Goodman, M. Ibe, A. Rajaraman, W. Shepherd, T. M. Tait, *et. al.*, *Gamma Ray Line Constraints on Effective Theories of Dark Matter*, *Nucl.Phys.* **B844** (2011) 55–68, [arXiv:1009.0008].

[94] K. Blum, M. Cliche, C. Csaki, and S. J. Lee, *WIMP Dark Matter through the Dilaton Portal*, *JHEP* **1503** (2015) 099, [arXiv:1410.1873].

[95] J. Callan, Curtis G., S. R. Coleman, and R. Jackiw, *A New improved energy - momentum tensor*, *Annals Phys.* **59** (1970) 42–73.

[96] C. Csaki, M. Graesser, L. Randall, and J. Terning, *Cosmology of brane models with radion stabilization*, *Phys.Rev.* **D62** (2000) 045015, [hep-ph/9911406].

[97] G. F. Giudice, R. Rattazzi, and J. D. Wells, *Graviscalars from higher dimensional metrics and curvature Higgs mixing*, *Nucl.Phys.* **B595** (2001) 250–276, [hep-ph/0002178].

[98] C. Csaki, M. L. Graesser, and G. D. Kribs, *Radion dynamics and electroweak physics*, *Phys.Rev.* **D63** (2001) 065002, [hep-th/0008151].

[99] C. Csaki, J. Hubisz, and S. J. Lee, *Radion phenomenology in realistic warped space models*, *Phys.Rev.* **D76** (2007) 125015, [arXiv:0705.3844].

[100] Y. Eshel, S. J. Lee, G. Perez, and Y. Soreq, *Shining Flavor and Radion Phenomenology in Warped Extra Dimension*, *JHEP* **1110** (2011) 015, [arXiv:1106.6218].

[101] **LUX Collaboration** Collaboration, D. Akerib *et. al.*, *First results from the LUX dark matter experiment at the Sanford Underground Research Facility*, *Phys.Rev.Lett.* **112** (2014) 091303, [arXiv:1310.8214].

[102] **XENON100 Collaboration** Collaboration, E. Aprile *et. al.*, *Dark Matter Results from 225 Live Days of XENON100 Data*, *Phys.Rev.Lett.* **109** (2012) 181301, [arXiv:1207.5988].

[103] **CDMS Collaboration** Collaboration, R. Agnese *et. al.*, *Silicon Detector Dark Matter Results from the Final Exposure of CDMS II*, *Phys.Rev.Lett.* **111** (2013) 251301, [arXiv:1304.4279].

[104] Y. Bai, M. Carena, and J. Lykken, *Dilaton-assisted Dark Matter*, *Phys.Rev.Lett.* **103** (2009) 261803, [arXiv:0909.1319].

[105] K. Agashe, K. Blum, S. J. Lee, and G. Perez, *Astrophysical Implications of a Visible Dark Matter Sector from a Custodially Warped-GUT*, *Phys.Rev.* **D81** (2010) 075012, [arXiv:0912.3070].

[106] Z. Chacko and R. K. Mishra, *Effective Theory of a Light Dilaton*, *Phys.Rev.* **D87** (2013), no. 11 115006, [arXiv:1209.3022].

[107] Z. Chacko, R. K. Mishra, and D. Stolarski, *Dynamics of a Stabilized Radion and Duality*, *JHEP* **1309** (2013) 121, [arXiv:1304.1795].

[108] B. Bellazzini, C. Csaki, J. Hubisz, J. Serra, and J. Terning, *A Naturally Light Dilaton and a Small Cosmological Constant*, *Eur.Phys.J.* **C74** (2014) 2790, [arXiv:1305.3919].

[109] F. Coradeschi, P. Lodone, D. Pappadopulo, R. Rattazzi, and L. Vitale, *A naturally light dilaton*, *JHEP* **1311** (2013) 057, [arXiv:1306.4601].

[110] P. J. Fox, R. Harnik, J. Kopp, and Y. Tsai, *Missing Energy Signatures of Dark Matter at the LHC*, *Phys.Rev.* **D85** (2012) 056011, [arXiv:1109.4398].

[111] P. Junnarkar and A. Walker-Loud, *Scalar strange content of the nucleon from lattice QCD*, *Phys.Rev.* **D87** (2013), no. 11 114510, [arXiv:1301.1114].

[112] Y. Bai, P. J. Fox, and R. Harnik, *The Tevatron at the Frontier of Dark Matter Direct Detection*, *JHEP* **1012** (2010) 048, [arXiv:1005.3797].

[113] J. R. Ellis, K. A. Olive, and C. Savage, *Hadronic Uncertainties in the Elastic Scattering of Supersymmetric Dark Matter*, *Phys.Rev.* **D77** (2008) 065026, [arXiv:0801.3656].

[114] K. Blum, *Cosmic ray propagation time scales: lessons from radioactive nuclei and positron data*, *JCAP* **1111** (2011) 037, [arXiv:1010.2836].

[115] J. Hisano, S. Matsumoto, and M. M. Nojiri, *Explosive dark matter annihilation*, *Phys.Rev.Lett.* **92** (2004) 031303, [hep-ph/0307216].

[116] T. R. Slatyer, *The Sommerfeld enhancement for dark matter with an excited state*, *JCAP* **1002** (2010) 028, [arXiv:0910.5713].

[117] A. Mayorov, O. Adriani, G. Barbarino, G. Bazilevskaia, R. Belotti, *et. al.*, *Antiprotons of galactic cosmic radiation in the PAMELA experiment*, *Bull.Russ.Acad.Sci.Phys.* **77** (2013) 602–605.

[118] B. Katz, K. Blum, and E. Waxman, *What can we really learn from positron flux 'anomalies'?*, *Mon.Not.Roy.Astron.Soc.* **405** (2010) 1458, [arXiv:0907.1686].

[119] M. Cirelli, G. Corcella, A. Hektor, G. Hutsi, M. Kadastik, *et. al.*, *PPPC 4 DM ID: A Poor Particle Physicist Cookbook for Dark Matter Indirect Detection*, *JCAP* **1103** (2011) 051, [arXiv:1012.4515].

[120] A. Kounine, *The Alpha Magnetic Spectrometer on the International Space Station*, *Int.J.Mod.Phys.* **E21** (2012), no. 08 1230005.

[121] V. Ginzburg, V. Dogiel, V. Berezinsky, S. Bulanov, and V. Ptuskin, *Astrophysics of cosmic rays, Amsterdam, Netherlands: North-Holland (1990) 534 p*, .

[122] **Fermi-LAT Collaboration** Collaboration, M. Ackermann *et. al.*, *Dark matter constraints from observations of 25 Milky Way satellite galaxies with the Fermi Large Area Telescope*, *Phys.Rev.* **D89** (2014), no. 4 042001, [arXiv:1310.0828].

[123] **H.E.S.S.Collaboration** Collaboration, A. Abramowski *et. al.*, *Search for a Dark Matter annihilation signal from the Galactic Center halo with H.E.S.S*, *Phys.Rev.Lett.* **106** (2011) 161301, [arXiv:1103.3266].

[124] J. F. Navarro, C. S. Frenk, and S. D. White, *The Structure of cold dark matter halos*, *Astrophys.J.* **462** (1996) 563–575, [astro-ph/9508025].

[125] **H.E.S.S. Collaboration** Collaboration, A. Abramowski *et. al.*, *Search for photon line-like signatures from Dark Matter annihilations with H.E.S.S*, *Phys.Rev.Lett.* **110** (2013) 041301, [arXiv:1301.1173].

[126] T. Cohen, M. Lisanti, A. Pierce, and T. R. Slatyer, *Wino Dark Matter Under Siege*, *JCAP* **1310** (2013) 061, [arXiv:1307.4082].

[127] J. Fan and M. Reece, *In Wino Veritas? Indirect Searches Shed Light on Neutralino Dark Matter*, *JHEP* **1310** (2013) 124, [arXiv:1307.4400].

[128] C. Dvorkin, K. Blum, and M. Kamionkowski, *Constraining Dark Matter-Baryon Scattering with Linear Cosmology*, *Phys.Rev.* **D89** (2014), no. 2 023519, [arXiv:1311.2937].

[129] S. Funk, *Indirect Detection of Dark Matter with gamma rays*, arXiv:1310.2695.

[130] M. Doro, *A decade of dark matter searches with ground-based Cherenkov telescopes*, *Nucl.Instrum.Meth.* **A742** (2014) 99–106, [arXiv:1404.5017].

[131] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, arXiv:1501.0596.

[132] G. Perez, *Top quark theory and the new physics searches frontier*, *Phys.Scripta* **T158** (2013) 014008.

[133] K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez, and J. Virzi, *LHC Signals from Warped Extra Dimensions*, *Phys.Rev.* **D77** (2008) 015003, [hep-ph/0612015].

[134] B. Lillie, L. Randall, and L.-T. Wang, *The Bulk RS KK-gluon at the LHC*, *JHEP* **0709** (2007) 074, [hep-ph/0701166].

[135] M. Perelstein and A. Spray, *Four boosted tops from a Regge gluon*, *JHEP* **1109** (2011) 008, [arXiv:1106.2171].

[136] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, *Stop Reconstruction with Tagged Tops*, *JHEP* **1010** (2010) 078, [arXiv:1006.2833].

[137] J. Berger, M. Perelstein, M. Saelim, and A. Spray, *Boosted Tops from Gluino Decays*, arXiv:1111.6594.

[138] A. Azatov, M. Salvarezza, M. Son, and M. Spannowsky, *Boosting Top Partner Searches in Composite Higgs Models*, *Phys.Rev.* **D89** (2014) 075001, [arXiv:1308.6601].

[139] T. Flacke, J. H. Kim, S. J. Lee, and S. H. Lim, *Constraints on composite quark partners from Higgs searches*, *JHEP* **1405** (2014) 123, [arXiv:1312.5316].

[140] M. Backović, G. Perez, T. Flacke, and S. J. Lee, *LHC Top Partner Searches Beyond the 2 TeV Mass Region*, arXiv:1409.0409.

[141] M. Backović, T. Flacke, J. H. Kim, and S. J. Lee, *Boosted Event Topologies from TeV Scale Light Quark Composite Partners*, arXiv:1410.8131.

[142] B. Gripaios, T. Mller, M. Parker, and D. Sutherland, *Search Strategies for Top Partners in Composite Higgs models*, *JHEP* **1408** (2014) 171, [arXiv:1406.5957].

[143] J. Reuter and M. Tonini, *Top Partner Discovery in the $T \to tZ$ channel at the LHC*, arXiv:1409.6962.

[144] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, *et. al.*, *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J.Phys.* **G39** (2012) 063001, [arXiv:1201.0008].

[145] M. Jankowiak and A. J. Larkoski, *Jet Substructure Without Trees*, *JHEP* **1106** (2011) 057, [arXiv:1104.1646].

[146] J. Bjorken and S. J. Brodsky, *Statistical Model for electron-Positron Annihilation Into Hadrons*, *Phys.Rev.* **D1** (1970) 1416–1420.

[147] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, *JHEP* **0807** (2008) 092, [arXiv:0806.0023].

[148] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, *et. al.*, *Substructure of high-p_T Jets at the LHC*, *Phys.Rev.* **D79** (2009) 074017, [arXiv:0807.0234].

[149] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys.Rev.Lett.* **100** (2008) 242001, [arXiv:0802.2470].

[150] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, *JHEP* **1002** (2010) 084, [arXiv:0912.1342].

[151] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys.Rev.* **D81** (2010) 094023, [arXiv:0912.0033].

[152] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **1103** (2011) 015, [arXiv:1011.2268].

[153] L. G. Almeida, O. Erdogan, J. Juknevich, S. J. Lee, G. Perez, *et. al.*, *Three-particle templates for a boosted Higgs boson*, *Phys.Rev.* **D85** (2012) 114046, [arXiv:1112.1957].

[154] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, and I. Sung, *Template Overlap Method for Massive Jets*, *Phys.Rev.* **D82** (2010) 054034, [arXiv:1006.2035].

[155] M. Backović, J. Juknevich, and G. Perez, *Boosting the Standard Model Higgs Signal with the Template Overlap Method*, *JHEP* **1307** (2013) 114, [arXiv:1212.2977].

[156] M. Backović, O. Gabizon, J. Juknevich, G. Perez, and Y. Soreq, *Measuring boosted tops in semi-leptonic $t\bar{t}$ events for the standard model and beyond*, *JHEP* **1404** (2014) 176, [arXiv:1311.2962].

[157] **D0 Collaboration** Collaboration, V. Abazov *et. al.*, *A precision measurement of the mass of the top quark*, *Nature* **429** (2004) 638–642, [hep-ex/0406031].

[158] P. Artoisenet, V. Lemaitre, F. Maltoni, and O. Mattelaer, *Automation of the matrix element reweighting method*, *JHEP* **1012** (2010) 068, [arXiv:1007.3300].

[159] D. E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, *Phys.Rev.* **D84** (2011) 074002, [arXiv:1102.3480].

[160] D. E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, arXiv:1211.3140.

[161] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft Drop*, *JHEP* **1405** (2014) 146, [arXiv:1402.2657].

[162] **ATLAS Collaboration** Collaboration, G. Aad *et. al.*, *A search for $t\bar{t}$ resonances in lepton+jets events with highly boosted top quarks collected in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **1209** (2012) 041, [arXiv:1207.2409].

[163] **ATLAS Collaboration** Collaboration, G. Aad *et. al.*, *Search for resonances decaying into top-quark pairs using fully hadronic decays in pp collisions with ATLAS at sqrt(s) = 7 TeV*, arXiv:1211.2202.

[164] **CMS Collaboration** Collaboration, *A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging, CMS-PAS-JME-09-001*, .

[165] **CMS Collaboration** Collaboration, *Jet Substructure Algorithms, CMS-PAS-JME-10-013*, .

[166] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **0302** (2003) 027, [hep-ph/0208156].

[167] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **0605** (2006) 026, [hep-ph/0603175].

[168] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput.Phys.Commun.* **178** (2008) 852–867, [arXiv:0710.3820].

[169] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur.Phys.J.* **C72** (2012) 1896, [arXiv:1111.6097].

[170] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **0804** (2008) 063, [arXiv:0802.1189].

[171] Y. Freund and R. E. Schapire, *A short introduction to boosting*, in *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 1401–1406, Morgan Kaufmann, 1999.

[172] A. J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **1309** (2013) 137, [arXiv:1307.1699].

[173] A. Hook, E. Izaguirre, M. Lisanti, and J. G. Wacker, *High Multiplicity Searches at the LHC Using Jet Masses*, *Phys.Rev.* **D85** (2012) 055029, [arXiv:1202.0558].

[174] T. Cohen, E. Izaguirre, M. Lisanti, and H. K. Lou, *Jet Substructure by Accident*, *JHEP* **1303** (2013) 161, [arXiv:1212.1456].

[175] T. R. Slatyer, *Signatures of a new force in the dark matter sector*. PhD thesis, Harvard University, 2010.

[176] P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, and K. E. Williams, *HiggsBounds: Confronting Arbitrary Higgs Sectors with Exclusion Bounds from LEP and the Tevatron*, *Comput. Phys. Commun.* **181** (2010) 138–167, [arXiv:0811.4169].

[177] P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, and K. E. Williams, *HiggsBounds 2.0.0: Confronting Neutral and Charged Higgs Sector Predictions with Exclusion Bounds from LEP and the Tevatron*, *Comput. Phys. Commun.* **182** (2011) 2605–2631, [arXiv:1102.1898].

[178] P. Bechtle *et. al.*, *Recent Developments in HiggsBounds and a Preview of HiggsSignals*, *PoS* **CHARGED2012** (2012) 024, [arXiv:1301.2345].

[179] M. Backović and J. Juknevich, *TemplateTagger v1.0.0: A Template Matching Tool for Jet Substructure*, arXiv:1212.2978.

[180] **ATLAS Collaboration** Collaboration, G. Aad *et. al.*, *A search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.* **D88** (2013) 012004, [arXiv:1305.2756].